

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORLED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C07K 2/00		A1	(11) International Publication Number: WO 98/20023 (43) International Publication Date: 14 May 1998 (14.05.98)
(21) International Application Number: PCT/AU97/00729 (22) International Filing Date: 31 October 1997 (31.10.97) (30) Priority Data: PO 3384 1 November 1996 (01.11.96) AU PO 5117 14 February 1997 (14.02.97) AU (71) Applicant (for all designated States except US): THE WALTER AND ELIZA HALL INSTITUTE OF MEDICAL RESEARCH [AU/AU]; Royal Parade, Parkville, VIC 3052 (AU). (72) Inventors; and (75) Inventors/Applicants (for US only): HILTON, Douglas, J. [AU/AU]; 244 Research Road, Warrandyte, VIC 3113 (AU). ALEXANDER, Warren, S. [AU/AU]; 13 Park Street, Moonee Ponds, VIC 3039 (AU). VINEY, Elizabeth, M. [AU/AU]; 7 Wallara Crescent, Bundoora, VIC 3083 (AU). WILLSON, Tracy, A. [AU/AU]; 26 Fortuna Avenue, North Balwyn, VIC 3104 (AU). RICHARDSON, Rachael, T. [AU/AU]; 24 Robinson Street, East Brighton, VIC 3187 (AU). STARR, Robyn [AU/AU]; 280 Drummond Street, Carlton, VIC 3053 (AU). NICHOLSON, Sandra, E. [AU/AU]; 66 Alma Terrace, Newport, VIC 3015 (AU). METCALF, Donald [AU/AU]; 268 Union Road,			(74) Agents: HUGHES, E., John, L. et al.; Davies Collison Cave, 1 Little Collins Street, Melbourne, VIC 3000 (AU). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i>
(54) Title: THERAPEUTIC AND DIAGNOSTIC AGENTS CAPABLE OF MODULATING CELLULAR RESPONSIVENESS TO CYTOKINES (57) Abstract <p>The present invention relates generally to therapeutic and diagnostic agents. More particularly, the present invention provides therapeutic molecules capable of modulating signal transduction such as but not limited to cytokine-mediated signal transduction. The molecules of the present invention are useful, therefore, in modulating cellular responsiveness to cytokines as well as other mediators of signal transduction such as endogenous or exogenous molecules, antigens, microbes and microbial products, viruses or components thereof, ions, hormones and parasites.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

- 1 -

THERAPEUTIC AND DIAGNOSTIC AGENTS CAPABLE OF MODULATING CELLULAR RESPONSIVENESS TO CYTOKINES**FIELD OF THE INVENTION**

5 The present invention relates generally to therapeutic and diagnostic agents. More particularly, the present invention provides therapeutic molecules capable of modulating signal transduction such as but not limited to cytokine-mediated signal transduction. The molecules of the present invention are useful, therefore, in modulating cellular responsiveness to cytokines as well as other mediators of signal transduction such as endogenous or exogenous molecules, antigens, microbes
10 and microbial products, viruses or components thereof, ions, hormones and parasites.

Bibliographic details of the publications referred to in this specification by author are collected at the end of the description. Sequence Identity Numbers (SEQ ID NOs.) for the nucleotide and amino acid sequences referred to in the specification are defined after the bibliography. A
15 summary of the SEQ ID NOs is given in Table 1.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other
20 integer or group of integers.

BACKGROUND OF THE INVENTION

Cells continually monitor their environment in order to modulate physiological and biochemical
25 processes which in turn affects future behaviour. Frequently, a cell's initial interaction with its surroundings occurs *via* receptors expressed on the plasma membrane. Activation of these receptors, whether through binding endogenous ligands (such as cytokines) or exogenous ligands (such as antigens), triggers a biochemical cascade from the membrane through the cytoplasm to the nucleus.

30

Of the endogenous ligands, cytokines represent a particularly important and versatile group.

- 2 -

Cytokines are proteins which regulate the survival, proliferation, differentiation and function of a variety of cells within the body [Nicola, 1994]. The haemopoietic cytokines have in common a four-alpha helical bundle structure and the vast majority interact with a structurally related family of cell surface receptors, the type I and type II cytokine receptors [Bazan, 1990; Sprang, 5 1993]. In all cases, ligand-induced receptor aggregation appears to be a critical event in initiating intracellular signal transduction cascades. Some cytokines, for example growth hormone, erythropoietin (Epo) and granulocyte-colony-stimulating factor (G-CSF), trigger receptor homodimerisation, while for other cytokines, receptor heterodimerisation or heterotrimerisation is crucial. In the latter cases, several cytokines share common receptor subunits and on this basis 10 can be grouped into three subfamilies with similar patterns of intracellular activation and similar biological effects [Hilton, 1994]. Interleukin-3 (IL-3), IL-5 and granulocyte-macrophage colony-stimulating factor (GM-CSF) use the common β -receptor subunit (βc) and each cytokine stimulates the production and functional activity of granulocytes and macrophages. IL-2, IL-4, IL-7, IL-9, and IL-15 each use the common γ -chain (γc), while IL-4 and IL-13 share an 15 alternative γ -chain ($\gamma'c$ or IL-13 receptor α -chain). Each of these cytokines plays an important role in regulating acquired immunity in the lymphoid system. Finally, IL-6, IL-11, leukaemia inhibitory factor (LIF), oncostatin-M (OSM), ciliary neurotrophic factor (CNTF) and cardiotrophin (CT) share the receptor subunit gp130. Each of these cytokines appears to be highly pleiotropic, having effects both within and outside the haemopoietic system [Nicola, 20 1994].

In all of the above cases at least one subunit of each receptor complex contains the conserved sequence elements, termed box1 and box2, in their cytoplasmic tails [Murakami, 1991]. Box1 is a proline-rich motif which is located more proximal to the transmembrane domain than the 25 acidic box 2 element. The box-1 region serves as the binding site for a class of cytoplasmic tyrosine kinases termed JAKs (Janus kinases). Ligand-induced receptor dimerisation serves to increase the catalytic activity of the associated JAKs through cross-phosphorylation. Activated JAKs then tyrosine phosphorylate several substrates, including the receptors themselves. Specific phosphotyrosine residues on the receptor then serve as docking sites for SH2-containing 30 proteins, the best characterised of which are the signal transducers and activators of transcription (STATs) and the adaptor protein, shc. The STATs are then phosphorylated on tyrosines,

- 3 -

probably by JAKs, dissociate from the receptor and form either homodimers or heterodimers through the interaction of the SH2 domain of one STAT with the phosphotyrosine residue of the other. STAT dimers then translocate to the nucleus where they bind to specific cytokine-responsive promoters and activate transcription [Darnell, 1994; Ihle, 1995; Ihle, 1995]. In a
5 separate pathway, tyrosine phosphorylated shc interacts with another SH2 domain-containing protein, Grb-2, leading ultimately to activation of members of the MAP kinase family and in turn transcription factors such as fos and jun [Sato, 1993; Cutler, 1993]. These pathways are not unique to members of the cytokine receptor family since cytokines that bind receptor tyrosine kinases also being able to activate STATs and members of the MAP kinase family [David, 1996;
10 Leaman, 1996; Shual, 1993; Sato, 1993; Cutler, 1993].

Four members of the JAK family of cytoplasmic tyrosine kinases have been described, JAK1, JAK2, JAK3 and TYK2, each of which binds to a specific subset of cytokine receptor subunits. Six STATs have been described (STAT1 through STAT6), and these too are activated by
15 distinct cytokine/receptor complexes. For example, STAT1 appears to be functionally specific to the interferon system, STAT4 appears to be specific to IL-12, while STAT6 appears to be specific for IL-4 and IL-13. Thus, despite common activation mechanisms some degree of cytokine specificity may be achieved through the use of specific JAKs and STATs [Thierfelder, 1996; Kaplan, 1996; Takeda, 1996; Shimoda, 1996; Meraz, 1996; Durbin, 1996].

20

In addition to those described above, there are clearly other mechanisms of activation of these pathways. For example, the JAK/STAT pathway appears to be able to activate MAP kinases independent of the shc-induced pathway [David, 1995] and the STATs themselves can be activated without binding to the receptor, possibly by direct interaction with JAKs [Gupta,
25 1996]. Conversely, full activation of STATS may require the action of MAP kinase in addition to that of JAKs [David, 1995; Wen, 1995].

While the activation of these signalling pathways is becoming better understood, little is known of the regulation of these pathways, including employment of negative or positive feedback
30 loops. This is important since once a cell has begun to respond to a stimulus, it is critical that the intensity and duration of the response is regulated and that signal transduction is switched

- 4 -

off. It is likewise desirable to increase the intensity of a response systemically or even locally as the situation requires.

In work leading up to the present invention, the inventors sought to isolate negative regulators of signal transduction. The inventors have now identified a new family of proteins which are capable of acting as regulators of signalling. The new family of proteins is defined as the suppressor of cytokine signalling (SOCS) family based on the ability of the initially identified SOCS molecules to suppress cytokine-mediated signalling. It should be noted, however, that not all members of the SOCS family need necessarily share suppressor function nor target solely cytokine mediated signalling. The SOCS family comprises at least three classes of protein molecules based on amino acid sequence motifs located N-terminal of a C-terminal motif called the SOCS box. The identification of this new family of regulatory molecules permits the generation of a range of effector or modulator molecules capable of modulating signal transduction and, hence, cellular responsiveness to a range of molecules including cytokines. The present invention, therefore, provides therapeutic and diagnostic agents based on SOCS proteins, derivatives, homologues, analogues and mimetics thereof as well as agonists and antagonists of SOCS proteins.

SUMMARY OF THE INVENTION

20

The present invention provides *inter alia* nucleic acid molecules encoding members of the SOCS family of proteins as well as the proteins themselves. Reference hereinafter to "SOCS" encompasses any or all members of the SOCS family. Specific SOCS molecules are defined numerically such as, for example, SOCS1, SOCS2 and SOCS3. The species from which the SOCS has been obtained may be indicated by a preface of a single letter abbreviation where "h" is human, "m" is murine and "r" is rat. Accordingly, "mSOCS1" is a specific SOCS from a murine animal. Reference herein to "SOCS" is not to imply that the protein solely suppresses cytokine-mediated signal transduction, as the molecule may modulate other effector-mediated signal transductions such as by hormones or other endogenous or exogenous molecules, antigens, microbes and microbial products, viruses or components thereof, ions, hormones and parasites. The term "modulates" encompasses up-regulation, down-regulation as well as

30

- 5 -

maintenance of particular levels.

One aspect of the present invention provides a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a protein or a derivative,
5 homologue, analogue or mimetic thereof or a nucleotide sequence capable of hybridizing thereto under low stringency conditions at 42°C wherein said protein comprises a SOCS box in its C-terminal region

Another aspect of the present invention provides a nucleic acid molecule comprising a sequence
10 of nucleotides encoding or complementary to a sequence encoding a protein or a derivative, homologue, analogue or mimetic thereof or a nucleotide sequence capable of hybridizing thereto under low stringency conditions at 42°C wherein said protein comprises a SOCS box in its C-terminal region and a protein:molecule interacting region.

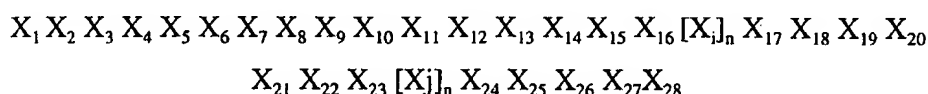
15 Yet another aspect of the present invention is directed to a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a protein or a derivative, homologue, analogue or mimetic thereof or a nucleotide sequence capable of hybridizing thereto under low stringency conditions at 42°C wherein said protein comprises a C-terminal region and a protein:molecule interacting region located in a region N-terminal of the
20 SOCS box.

Preferably, the protein:molecule interacting region is a protein:DNA or protein:protein binding region.

25 Still a further aspect of the present invention provides a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a protein or a derivative, homologue, analogue or mimetic thereof or a nucleotide sequence capable of hybridizing thereto under low stringency conditions at 42°C wherein said protein comprises a SOCS box in its C-terminal region and one or more of an SH2 domain, WD-40 repeats or
30 ankyrin repeats N-terminal of the SOCS box.

- 6 -

Even still a further aspect of the present invention is directed to a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a protein or a derivative, homologue, analogue or mimetic thereof or a nucleotide sequence capable of hybridizing thereto under low stringency conditions at 42°C wherein said protein
 5 comprises a SOCS box in its C-terminal region wherein the SOCS box comprises the amino acid sequence:



10

wherein: X_1 is L, I, V, M, A or P;
 X_2 is any amino acid residue;
 X_3 is P, T or S;
 X_4 is L, I, V, M, A or P;
 X_5 is any amino acid;
 X_6 is any amino acid;
 X_7 is L, I, V, M, A, F, Y or W;
 X_8 is C, T or S;
 X_9 is R, K or H;

15

20

X_{10} is any amino acid;
 X_{11} is any amino acid;
 X_{12} is L, I, V, M, A or P;
 X_{13} is any amino acid;

25

X_{14} is any amino acid;
 X_{15} is any amino acid;

X_{16} is L, I, V, M, A, P, G, C, T or S;

$[X_i]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;

30

X_{17} is L, I, V, M, A or P;
 X_{18} is any amino acid;

- 7 -

X_{19} is any amino acid;

X_{20} L, I, V, M, A or P;

X_{21} is P;

X_{22} is L, I, V, M, A, P or G;

5 X_{23} is P or N;

$[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_j may comprise the same or different amino acids selected from any amino acid residue;

X_{24} is L, I, V, M, A or P;

10 X_{25} is any amino acid;

X_{26} is any amino acid;

X_{27} is Y or F;

X_{28} is L, I, V, M, A or P;

15 and a protein:molecule interacting region such as but not limited to one or more of an SH2 domain, WD-40 repeats and/or ankyrin repeats N-terminal of the SOCS box.

Another aspect of the present invention is directed to a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a protein or a
20 derivative, homologue, analogue or mimetic thereof or a nucleotide sequence capable of hybridizing thereto under low stringency conditions at 42°C wherein said protein exhibits the following characteristics:

(i) comprises a SOCS box in its C-terminal region having the amino acid sequence:

25 $X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14} X_{15} X_{16} [X]_n X_{17} X_{18} X_{19} X_{20}$
 $X_{21} X_{22} X_{23} [X]_n X_{24} X_{25} X_{26} X_{27} X_{28}$

wherein: X_1 is L, I, V, M, A or P;

X_2 is any amino acid residue;

30 X_3 is P, T or S;

X_4 is L, I, V, M, A or P;

- 8 -

X_5 is any amino acid;

X_6 is any amino acid;

X_7 is L, I, V, M, A, F, Y or W;

X_8 is C, T or S;

5 X_9 is R, K or H;

X_{10} is any amino acid;

X_{11} is any amino acid;

X_{12} is L, I, V, M, A or P;

X_{13} is any amino acid;

10 X_{14} is any amino acid;

X_{15} is any amino acid;

X_{16} is L, I, V, M, A, P, G, C, T or S;

$[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids

and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;

15 X_{17} is L, I, V, M, A or P;

X_{18} is any amino acid;

X_{19} is any amino acid;

X_{20} L, I, V, M, A or P;

20 X_{21} is P;

X_{22} is L, I, V, M, A, P or G;

X_{23} is P or N;

$[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids

and wherein the sequence X_j may comprise the same or different amino acids selected from any amino acid residue;

25 X_{24} is L, I, V, M, A or P;

X_{25} is any amino acid;

X_{26} is any amino acid;

X_{27} is Y or F;

30 X_{28} is L, I, V, M, A or P; and

(ii) comprises at least one of a SH2 domain, WD-40 repeats and/or ankyrin repeats or other

- 9 -

protein:molecule interacting domain in a region N-terminal of the SOCS box.

Preferably, the SOCS molecules modulate signal transduction such as from a cytokine or hormone or other endogenous or exogenous molecule, a microbe or microbial product, an antigen or a parasite.

More preferably, the SOCS molecule modulate cytokine mediated signal transduction.

Still another aspect of the present invention comprises a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a protein or a derivative, homologue, analogue or mimetic thereof or comprises a nucleotide sequence capable of hybridizing thereto under low stringency conditions at 42°C wherein said protein exhibits the following characteristics;

- (i) is capable of modulating signal transduction;
- (ii) comprises a SOCS box in its C-terminal region having the amino acid sequence:

$$X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14} X_{15} X_{16} [X_i]_n X_{17} X_{18} X_{19} X_{20} \\ X_{21} X_{22} X_{23} [X_j]_n X_{24} X_{25} X_{26} X_{27} X_{28}$$

- wherein:
- X₁ is L, I, V, M, A or P;
 - X₂ is any amino acid residue;
 - X₃ is P, T or S;
 - X₄ is L, I, V, M, A or P;
 - X₅ is any amino acid;
 - X₆ is any amino acid;
 - X₇ is L, I, V, M, A, F, Y or W;
 - X₈ is C, T or S;
 - X₉ is R, K or H;
 - X₁₀ is any amino acid;
 - X₁₁ is any amino acid;
 - X₁₂ is L, I, V, M, A or P;

- 10 -

X_{13} is any amino acid;

X_{14} is any amino acid;

X_{15} is any amino acid;

X_{16} is L, I, V, M, A, P, G, C, T or S;

5 $[X_i]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;

X_{17} is L, I, V, M, A or P;

X_{18} is any amino acid;

10 X_{19} is any amino acid;

X_{20} L, I, V, M, A or P;

X_{21} is P;

X_{22} is L, I, V, M, A, P or G;

X_{23} is P or N;

15 $[X_j]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_j may comprise the same or different amino acids selected from any amino acid residue;

X_{24} is L, I, V, M, A or P;

X_{25} is any amino acid;

20 X_{26} is any amino acid;

X_{27} is Y or F;

X_{28} is L, I, V, M, A or P; and

(iii) comprises at least one of a SH2 domain, WD-40 repeats and/or ankyrin repeats or other
25 protein:molecule interacting domain in a region N-terminal of the SOCS box.

Preferably, the signal transduction is mediated by a cytokine such as one or more of EPO, TPO, G-CSF, GM-CSF, IL-3, IL-2, IL-4, IL-7, IL-13, IL-6, LIF, IL-12, IFN α , TNF α , IL-1 and/or M-CSF.

30

Preferably, the signal transduction is mediated by one or more of Interleukin 6 (IL-6), Leukaemia

- 11 -

Inhibitory Factor (LIF), Oncostatin M (OSM), Interferon (IFN)- α and/or thrombopoietin.

Preferably, the signal transduction is mediated by IL-6.

5 Particularly preferred nucleic acid molecules comprise nucleotide sequences substantially set forth in SEQ ID NO:3 (mSOCS1), SEQ ID NO:5 (mSOCS2), SEQ ID NO:7 (mSOCS3), SEQ ID NO:9 (hSOCS1), SEQ ID NO:11 (rSOCS1), SEQ ID NO:13 (mSOCS4), SEQ ID NO:15 and SEQ ID NO:16 (hSOCS4), SEQ ID NO:17 (mSOCS5), SEQ ID NO:19 (hSOCS5), SEQ ID NO:20 (mSOCS6), SEQ ID NO:22 and SEQ ID NO:23 (hSOCS6), SEQ ID NO:24
10 (mSOCS7), SEQ ID NO:26 and SEQ ID NO:27 (hSOCS7), SEQ ID NO:28 (mSOCS8), SEQ ID NO:30 (mSOCS9), SEQ ID NO:31 (hSOCS9), SEQ ID NO:32 (mSOCS10), SEQ ID NO:33 and SEQ ID NO:34 (hSOCS10), SEQ ID NO:35 (hSOCS11), SEQ ID NO:37 (mSOCS12), SEQ ID NO:38 and SEQ ID NO:39 (hSOCS12), SEQ ID NO:40 (mSOCS13), SEQ ID NO:42 (hSOCS13), SEQ ID NO: 43 (mSOCS14), SEQ ID NO:45 (mSOCS15) and SEQ ID NO:47
15 (hSOCS15) or a nucleotide sequence having at least about 15% similarity to all or a region of any of the listed sequences or a nucleotide acid molecule capable of hybridizing to any one of the listed sequences under low stringency conditions at 42°C.

Another aspect of the present invention relates to a protein or a derivative, homologue, analogue
20 or mimetic thereof comprising a SOCS box in its C-terminal region.

Yet another aspect of the present invention is directed to a protein or a derivative, homologue, analogue or mimetic thereof comprising a SOCS box in its C-terminal region and a protein:molecule interacting region.

25

Even yet another aspect of the present invention provides a protein or a derivative, homologue, analogue or mimetic thereof comprising an interacting region located in a region N-terminal of the SOCS box.

30 Preferably, the protein:molecule interacting region is a protein:DNA or a protein:protein binding region.

- 12 -

Another aspect of the present invention contemplates a protein or a derivative, homologue, analogue or mimetic thereof comprising a SOCS box in its C-terminal region and a SH2 domain, WD-40 repeats or ankyrin repeats N-terminal of the SOCS box.

5 Still yet another aspect of the present invention provides a protein or a derivative, homologue, analogue or mimetic thereof exhibiting the following characteristics:

(i) comprises a SOCS box in its C-terminal region having the amino acid sequence:

10 $X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14} X_{15} X_{16} [X_i]_n X_{17} X_{18} X_{19} X_{20}$
 $X_{21} X_{22} X_{23} [X_j]_n X_{24} X_{25} X_{26} X_{27} X_{28}$

wherein: X_1 is L, I, V, M, A or P;

X_2 is any amino acid residue;

15 X_3 is P, T or S;

X_4 is L, I, V, M, A or P;

X_5 is any amino acid;

X_6 is any amino acid;

X_7 is L, I, V, M, A, F, Y or W;

20 X_8 is C, T or S;

X_9 is R, K or H;

X_{10} is any amino acid;

X_{11} is any amino acid;

X_{12} is L, I, V, M, A or P;

25 X_{13} is any amino acid;

X_{14} is any amino acid;

X_{15} is any amino acid;

X_{16} is L, I, V, M, A, P, G, C, T or S;

$[X_i]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids

30 and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;

- 13 -

X_{17} is L, I, V, M, A or P;

X_{18} is any amino acid;

X_{19} is any amino acid;

X_{20} L, I, V, M, A or P;

5 X_{21} is P;

X_{22} is L, I, V, M, A, P or G;

X_{23} is P or N;

$[X_j]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_j may comprise the same or different amino acids selected from any amino acid residue;

10

X_{24} is L, I, V, M, A or P;

X_{25} is any amino acid;

X_{26} is any amino acid;

X_{27} is Y or F;

15

X_{28} is L, I, V, M, A or P; and

(ii) comprises at least one of a SH2 domain, WD-40 repeats and/or ankyrin repeats or other protein:molecule interacting domain in a region N-terminal of the SOCS box.

20 Preferably, the proteins modulate signal transduction such as cytokine-mediated signal transduction.

Preferred cytokines are EPO, TPO, G-CSF, GM-CSF, IL-3, IL-2, IL-4, IL-7, IL-13, IL-6, LIF, IL-12, IFN γ , TNF α , IL-1 and/or M-CSF.

25

A particularly preferred cytokine is IL-6.

Even yet another aspect of the present invention provides a protein or derivative, homologue, analogue or mimetic thereof exhibiting the following characteristics:

30 (i) is capable of modulating signal transduction such as cytokine-mediated signal transduction;

- 14 -

- (ii) comprises a SOCS box in its C-terminal region having the amino acid sequence:

$$X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14} X_{15} X_{16} [X_i]_n X_{17} X_{18} X_{19} X_{20} \\ X_{21} X_{22} X_{23} [X_j]_n X_{24} X_{25} X_{26} X_{27} X_{28}$$

5

wherein: X_1 is L, I, V, M, A or P;

X_2 is any amino acid residue;

X_3 is P, T or S;

X_4 is L, I, V, M, A or P;

10

X_5 is any amino acid;

X_6 is any amino acid;

X_7 is L, I, V, M, A, F, Y or W;

X_8 is C, T or S;

X_9 is R, K or H;

15

X_{10} is any amino acid;

X_{11} is any amino acid;

X_{12} is L, I, V, M, A or P;

X_{13} is any amino acid;

X_{14} is any amino acid;

20

X_{15} is any amino acid;

X_{16} is L, I, V, M, A, P, G, C, T or S;

$[X_i]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;

25

X_{17} is L, I, V, M, A or P;

X_{18} is any amino acid;

X_{19} is any amino acid;

X_{20} is L, I, V, M, A or P;

X_{21} is P;

30

X_{22} is L, I, V, M, A, P or G;

X_{23} is P or N;

- 15 -

$[X_i]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;

X_{24} is L, I, V, M, A or P;

5 X_{25} is any amino acid;

X_{26} is any amino acid;

X_{27} is Y or F;

X_{28} is L, I, V, M, A or P; and

- 10 (iii) comprises at least one of a SH2 domain, WD-40 repeats and/or ankyrin repeats or other protein-molecule interacting domain in a region N-terminal of the SOCS box.

Particularly preferred SOCS proteins comprise an amino acid sequence substantially as set forth in SEQ ID NO:4 (mSOCS1), SEQ ID NO:6 (mSOCS2), SEQ ID NO:8 (mSOCS3), SEQ ID
15 NO:10 (hSOCS1), SEQ ID NO:12 (rSOCS1), SEQ ID NO:14 (mSOCS4), SEQ ID NO:18 (mSOCS5), SEQ ID NO:21 (mSOCS6), SEQ ID NO:25 (mSOCS7), SEQ ID NO:29 (mSOCS8), SEQ ID NO:36 (hSOCS11), SEQ ID NO:41 (mSOCS13), SEQ ID NO:44 (mSOCS14), SEQ ID NO:46 (mSOCS15) and SEQ ID NO:48 (hSOCS15) or an amino acid sequence having at least 15% similarity to all or a region of any one of the listed sequences.

20

Another aspect of the present invention contemplates a method of modulating levels of a SOCS protein in a cell said method comprising contacting a cell containing a SOCS gene with an effective amount of a modulator of SOCS gene expression or SOCS protein activity for a time and under conditions sufficient to modulate levels of said SOCS protein.

25

A related aspect of the present invention provides a method of modulating signal transduction in a cell containing a SOCS gene comprising contacting said cell with an effective amount of a modulator of SOCS gene expression or SOCS protein activity for a time sufficient to modulate signal transduction.

30

Yet a further related aspect of the present invention is directed to a method of influencing

- 16 -

interaction between cells wherein at least one cell carries a SOCS gene, said method comprising contacting the cell carrying the SOCS gene with an effective amount of a modulator of SOCS gene expression or SOCS protein activity for a time sufficient to modulate signal transduction.

- 5 In accordance with the present invention, n in $[X_i]_n$ and $[X_j]_n$ may, in addition from being 1-50, be from 1-30, 1-20, 1-10 and 1-5.

A summary of the SEQ ID NOs referred to in the subject specification is given in Table 1.

10

- 17 -

TABLE 1
SUMMARY OF SEQUENCE IDENTITY NUMBERS

	SEQUENCE	SEQ ID NO.
5	PCR Primer	1
	PCR Primer	2
	Mouse SOCS1 (nucleotide)	3
	Mouse SOCS1 (amino acid)	4
10	Mouse SOCS2 (nucleotide)	5
	Mouse SOCS2 (amino acid)	6
	Mouse SOCS3 (nucleotide)	7
	Mouse SOCS3 (amino acid)	8
	Human SOCS1 (nucleotide)	9
15	Human SOCS1 (amino acid)	10
	Rat SOCS1 (nucleotide)	11
	Rat SOCS1 (amino acid)	12
	nucleotide sequence of murine SOCS4	13
	amino acid sequence of murine SOCS4	14
20	nucleotide sequence of SOCS4 cDNA human contig 4.1	15
	nucleotide sequence of SOCS4 cDNA human contig 4.2	16
	nucleotide sequence of murine SOCS5	17
	amino acid sequence of murine SOCS5	18
	nucleotide sequence of human SOCS5	19
25	nucleotide sequence of murine SOCS6	20
	amino acid of murine SOCS6	21
	nucleotide sequence of human SOCS6 contig h6.1	22
	nucleotide sequence of human SOCS6 contig h6.2	23
	nucleotide sequence of murine SOCS7	24

- 18 -

	amino acid sequence of murine SOCS7	25
	nucleotide sequence of human SOCS7 contig h7.1	26
	nucleotide sequence of human SOCS7 contig 17.2	27
	nucleotide sequence of murine SOCS8	28
5	amino acid sequence of murine SOCS 8	29
	nucleotide sequence of murine SOCS9	30
	nucleotide sequence of human SOCS9	31
	nucleotide sequence of murine SOCS10	32
	nucleotide sequence of human SOCS10 contig h10.1	33
10	nucleotide sequence of human SOCS10 contig h10.2	34
	nucleotide sequence of human SOCS11	35
	amino acid sequence of human SOCS11	36
	nucleotide sequence of mouse SOCS12	37
	nucleotide sequence of human SOCS12 contig h12.1	38
15	nucleotide sequence of human SOCS12 contig h12.2	39
	nucleotide sequence of murine SOCS13	40
	amino acid sequence of murine SOCS13	41
	nucleotide sequence of human SOCS13 cDNA contig h13.1	42
	nucleotide sequence of murine SOCS14 cDNA	43
20	amino acid sequence of murine SOCS14	44
	nucleotide sequence of murine SOCS15 cDNA	45
	amino acid sequence of murine SOCS15	46
	nucleotide sequence of human SOCS15	47
	amino acid sequence of human SOCS15	48

25

- 19 -

Single and three letter abbreviations are used to denote amino acid residues and these are summarized in Table 2.

TABLE 2

5	Amino Acid	Three-letter Abbreviation	One-letter Symbol
	Alanine	Ala	A
10	Arginine	Arg	R
	Asparagine	Asn	N
	Aspartic acid	Asp	D
	Cysteine	Cys	C
	Glutamine	Gln	Q
15	Glutamic acid	Glu	E
	Glycine	Gly	G
	Histidine	His	H
	Isoleucine	Ile	I
	Leucine	Leu	L
20	Lysine	Lys	K
	Methionine	Met	M
	Phenylalanine	Phe	F
	Proline	Pro	P
	Serine	Ser	S
25	Threonine	Thr	T
	Tryptophan	Trp	W
	Tyrosine	Tyr	Y
	Valine	Val	V
	Any residue	Xaa	X
30			

- 20 -

BRIEF DESCRIPTION OF THE DRAWINGS

In some of the Figures, abbreviations are used to denote SOCS proteins with certain binding motifs. SOCS proteins which contain WD-40 repeats are referred to as WSB1-WSB4. SOCS
5 proteins with ankyrin repeats are referred to as ASB1-ASB3.

Figure 1 is a diagrammatic representation showing generation of an IL-6-unresponsive M1 clone by retroviral infection. The RUFneo retrovirus, showing the position of landmark restriction endonuclease cleavage sites, the 4A2 cDNA insert and the position of PCR primer sequences.

10

Figure 2 is a photographic representation of Southern and Northern analysis. (Left and Middle Panels) Southern blot analysis of genomic DNA from clone 4A2 and a control infected M1 clone. DNA was digested with BamH I, to reveal the number of retroviruses carried by each clone, and Sac I, to estimate the size of the retroviral cDNA insert. Left panel; probed with neo. Right
15 panel; probed with the Xho I-digested 4A2 PCR product. (Right Panel) . Northern blot analysis of total RNA from clone 4A2 and a control infected M1 clone, probed with the Xho I-digested 4A2 PCR product. The two bands represent unspliced and spliced retroviral transcripts, resulting from splice donor and acceptor sites in the retroviral genome.

20 **Figure 3** is a representation of the nucleotide sequence and structure of the SOCS1 gene. A. The genomic context of SOCS1 in relation to the protamine gene cluster on murine chromosome 16. The accession number of this locus is MPMRGMNS (direct submission; G. Schlueter, 1995) for the mouse and BTPRMTNP2 for the rat (direct submission; G. Schlueter, 1996). B. The nucleotide sequence of the SOCS1 cDNA and deduced amino acid sequence. Conventional one
25 letter abbreviations are used for the amino acid sequence and the asterisk indicates the stop codon. The polyadenylation signal sequence is underlined. The coding region is shown in uppercase and the untranslated region is shown in lower case.

Figure 4 is a graphical representation of cell differentiation in the presence of cytokines. Semi-
30 solid agar cultures of parental M1 cells (M1 and M1.mpl) and M1 cells expressing SOCS1 (4A2 and M1.mpl.SOCS1), were used and the percentage of colonies which differentiated in response

- 21 -

to a titration of 1 mg/ml IL-6 (●), 100 ng/ml LIF (◇), 1 mg/ml OSM (□), 100 ng/ml IFN-γ (▲), 500 ng/ml TPO (●), or 3×10^{-6} M dexamethasone (*) determined.

Figure 5 is a photographic representation of cytopins of liquid cultures of parental M1 cells (M1 and M1.mpl) and M1 cells expressing SOCS1 (4A2 and M1.mpl.SOCS1) cultured for 4 days in the presence of 10 ng/ml IL-6 or saline. Unlike parental M1 cells, morphological features consistent with macrophage differentiation are not observed in M1 cells constitutively expressing SOCS1 (4A2 and M1.mpl.SOCS1) when cultured in IL-6.

Figure 6 is a photographic representation showing inhibition of phosphorylation of signalling molecules by SOCS1. Parental M1 cells (M1 and M1.mpl) and M1 cells expressing SOCS1 (4A2 and M1.mpl.SOCS1) were incubated in the absence (-) or presence (+) of 10 ng/ml of IL-6 for 4 minutes at 37°C. Cells were then lysed and extracts were either immunoprecipitated using anti-mouse gp130 antibody prior to SDS-PAGE (two upper panels) or were electrophoresed directly (two lower panels). Gels were blotted and the filters were then probed with anti-phosphotyrosine (upper panel), anti-gp130 antibody (second top panel), anti-phospho-STAT3 (second bottom panel) or anti-STAT3 (lower panel). Blots were visualised using peroxidase-conjugated secondary antibodies and Enhanced Chemiluminescence (ECL) reagents.

Figure 7 is a representation of protein extracts prepared from (A) M1 cells or M1 cells expressing SOCS1 (4A2) and (B) M1.mpl cells or M1.mpl.SOCS1 cells incubated for 10 min at 37°C in 10 ml serum-free DME containing either saline, 100 ng/ml IL-6 or 100 ng/ml IFN-γ. The binding reactions contained 4-6 µg protein (constant within a given experiment), 5 ng ³²P-labelled m67 oligonucleotide encoding the high affinity SIF (c-sis- inducible factor) binding site, and 800 ng sonicated salmon sperm DNA. For certain experiments, protein samples were preincubated with an excess of unlabelled m67 oligonucleotide, or antibodies specific for either STAT1 or STAT3.

Figure 8 is a photographic representation of Northern hybridisation. Mice were injected intravenously with 2 µg and after various periods of time, the livers were removed and polyA+

- 22 -

mRNA was purified. M1 cells were stimulated for various lengths of time with 500 ng/ml of IL-6, after which polyA+ mRNA was isolated. mRNA was fractionated by electrophoresis and immobilized on nylon filters. Northern blots were prehybridized, hybridized with random-primed ³²P-labelled SOCS1 or GAPDH DNA fragments, washed and exposed to film overnight.

5

Figure 9 is a representation of a comparison of the amino acid sequences of SOCS1, SOCS2, SOCS3 and CIS. Alignment of the predicted amino acid sequence of mouse (mm), human (hs) and rat (rr) SOCS1, SOCS2, SOCS3 and CIS. Those residues shaded are conserved in three or four mouse SOCS family members. The SH2 domain is boxed in solid lines, while the SOCS box
10 is bounded by double lines.

Figure 10 is a photographic representation showing the phenotype of IL-6 unresponsive M1 cell clone, 4A2. Colonies of parental M1 cells (left panel) and clone 4A2 (right panel) cultured in semi-solid agar for 7 days in saline or 100 ng/ml IL-6.

15

Figure 11 is a photographic representation showing expression of mRNA for SOCS family members *in vitro* and *in vivo*.

(A) Northern analysis of mRNA from a range of mouse organs showing constitutive expression of SOCS family members in a limited number of tissues.

20 (B) Northern analysis of mRNA from liver and M1 cells showing induction of expression of SOCS family members following exposure to IL-6.

(C) Reverse transcriptase PCR analysis of mRNA from bone marrow showing induction of expression of SOCS family members by a range of cytokines.

25 **Figure 12** is a photographic representation showing SOCS1 suppresses the phosphorylation and activation of gp130 and STAT-3.

(A) Western blots of extracts from parental M1 cells (M1 and M1.mpl) and M1 cells expressing SOCS1 (4A2 and M1.mpl.SOCS1) stimulated with (+) or without (-) 100 ng/ml IL-6.

Top: Extracts immunoprecipitated with anti-gp130 (αgp130) and immunoblotted with anti-phosphotyrosine (αPY-STAT3), or for STAT3 (αSTAT3) to demonstrate equal loading of
30 protein. The molecular weights of the bands are shown on the right.

- 23 -

(B) EMSA of M1.mpl and M1.mpl.SOCS1 cells stimulated with (+) and without (-) 100 ng/ml IL-6 or 100 ng/ml IFN γ . The DNA-binding complexes SIF A, B, and C are indicated at the left.

5 **Figure 13** is a representation of a comparison of the amino acid sequence of the SOCS proteins
 (A) Schematic representation of structures of SOCS proteins including proteins which contain
 WD-40 repeats (WSB) and ankyrin repeats (ASB). (B) Alignment of N-terminal regions of
 SOCS proteins. (C) Alignment of the SH2 domains of CIS, SOCS1, 2, 3, 5, 9, 11 and 14. (D)
 Alignment of the WD-40 repeats of SOCS4, SOCS6, SOCS13 and SOCS15. (E) Alignment of
 10 the ankyrin repeats of SOCS7 and SOCS10. (F) Alignment of the regions between SH2, WD-40
 and ankyrin repeats and the SOCS box. (G) Alignment of the SOCS box. In each case the
 conventional one letter abbreviations for amino acids are used, with X denoting residues of
 uncertain identity and $\circ\circ\circ$ denoting the beginning and the end of contigs. Amino acid
 sequence obtained from conceptual translation of nucleic acid sequence derived from isolated
 15 cDNAs is shown in upper case while amino acid sequence obtained by conceptual translation of
 ESTs is shown in lower case and is approximate only. Conserved residues, defined as (LIVMA),
 (FYW), (DE), (QN), (C, S, T), (KRH), (PG) are shaded in the SH2 domain, WD-40 repeats,
 ankyrin repeats and the SOCS box. For the alignment of SH2 domains, WD-40 repeats and
 ankyrin repeats a consensus sequence is shown above. In each case this has been derived from
 20 examination of a large and diverse set of domains (Neer *et al*, 1994; Bork, 1993).

Figures 14(A) and (B) are photographic representations showing analysis of mRNA expression
 of mouse SOCS1 and SOCS5 and SOCS containing a WD-40 repeat (WSB2) and ankyrin
 repeats (ASB1).

25

Figure 15 is a representation showing the nucleotide sequence of the mouse SOCS4 cDNA. The
 nucleotides encoding the mature coding region from the predicted ATG "start" codon to the stop
 codon is shown in upper case, while the predicted 5' and 3' untranslated regions are shown in
 lower case. The relationship of mouse cDNA sequence to mouse and human EST contigs is
 30 illustrated in Figure 17.

- 24 -

Figure 16 is a representation showing the predicted amino acid sequence of the mouse SOCS4 protein, derived from the nucleotide sequence in Figure 15. The SOCS box, which also shown in Figure 13, is underlined.

5 **Figure 18** is a representation showing the nucleotide sequence of human SOCS4 cDNA contigs h4.1 and h4.2, derived from analysis of ESTs listed in Table 4.1. The relationship of these contigs to the mouse cDNA sequence is illustrated in Figure 17.

Figure 19 is a diagrammatic representation showing the relationship of mouse SOCS5 genomic
10 (57-2) and cDNA (5-3-2) clones to contigs derived from analysis of mouse ESTs (Table 5.1) and human cDNA clone (5-94-2) and ESTs (Table 5.2). The nucleotide sequence of the mouse SOCS5 contig is shown in Figure 20, with the sequence of human SOCS5 contig (h5.1) being shown in Figure 21. The deduced amino acid sequence of mouse SOCS5 is shown in Figure 20B. The structure of the protein is shown schematically, with the SH2 domain indicated by
15 () and the SOCS box by (). The putative 5' and 3' translated regions are shown by the thin solid line.

Figure 20A is a representation showing the nucleotide sequence of the mouse SOCS5 derived from analysis of genomic and cDNA clones. The nucleotides encoding the mature coding region
20 from the predicted ATG "start" codon to the stop codon is shown in upper case, while the predicted 5' and 3' untranslated regions are shown in lower case. The relationship of mouse cDNA sequence to mouse and human EST contigs is illustrated in Figure 19.

Figure 20B is a representation of the predicted amino acid sequence of mouse SOCS5 protein,
25 derived from the nucleotide sequence in Figure 20A. The SOCS box, which also shown in Figure 13 is underlined.

Figure 21 is a representation showing the nucleotide sequence of human SOCS5 cDNA contig h5.1, derived from analysis of cDNA clone 5-94-2 and the ESTs listed in Table 5.2. The
30 relationship of these contigs to the mouse cDNA sequence is illustrated in Figure 19.

- 25 -

Figure 22 is a diagrammatic representation showing the relationship of mouse SOCS6 cDNA clones (6-1A, 6-2A, 6-5B, 6-4N, 6-18, 6-29, 6-3N and 6-5N) to contigs derived from analysis of mouse ESTs (Table 6.1) and human ESTs (Table 6.2). The nucleotide sequence of the mouse SOCS-6 contig is shown in Figure 23, with the sequence of human SOCS6 contigs (h6.1 and h6.2) being shown in Figure 24. The deduced amino acid sequence of mouse SOCS6 is shown in Figure 23B. The structure of the protein is shown schematically, while the WD-40 repeats indicated by () and the SOCS box by (). The putative 5' and 3' untranslated regions are shown by the thin solid line.

Figure 23A is a representation showing the nucleotide sequence of the mouse SOCS6 derived from analysis of cDNA clone 64-10A-11. The nucleotides encoding the part of the predicted coding region, ending in the stop codon are shown in upper case, while the predicted 3' untranslated regions are shown in lower case. The relationship of mouse cDNA sequence to mouse and human EST contigs is illustrated in Figure 22.

15

Figure 23B is a representation showing the predicted amino acid sequence of mouse SOCS6 protein, derived from the nucleotide sequence in Figure 23A. The SOCS box, which also shown in Figure 13 is underlined.

Figure 24 is a representation showing the nucleotide sequence of human SOCS6 cDNA contig h6.1, derived from analysis of cDNA clone 5-94-2 and the ESTs listed in Table 6.2. The relationship of these contigs to the mouse cDNA sequence is illustrated in Figure 22

Figure 25 is a diagrammatic representation showing the relationship of mouse SOCS7 cDNA clone (74-10A-11) to contigs derived from analysis of mouse ESTs (Table 7.1) and human ESTs (Table 7.2). The nucleotide sequence of the mouse SOCS7 contig is shown in Figure 26 with the sequence of human SOCS7 contigs (h7.1 and h7.2) being shown in Figure 27. The deduced amino acid sequence of mouse SOCS7 is shown in Figure 26B. The structure of the protein is shown schematically, with the ankyrin repeats indicated by () and the SOCS box by (). The putative 5' and 3' untranslated regions are shown by the thin solid line in the mouse and by the wavy line in h7.2. Based on analysis of clones isolated to date and ESTs the 3' untranslated

- 26 -

regions of mSOCS7 and hSOCS7 share little similarity.

Figure 26A is a representation showing the nucleotide sequence of the mouse SOCS7 derived from analysis of cDNA clone 74-10A-11. The nucleotides encoding the part of the predicted
5 coding region, ending in the stop codon are shown in upper case, while the predicted 3' untranslated regions are shown in lower case. The relationship of mouse cDNA sequence to mouse and human EST contigs is illustrated in Figure 25.

Figure 26B is a representation showing the predicted amino acid sequence of mouse SOCS7
10 protein, derived from the nucleotide sequence in Figure 26A. The SOCS box, which also shown in Figure 13 is underlined.

Figure 27 is a representation showing the nucleotide sequence of human SOCS7 cDNA contig h7.1 and h7.2 derived from analysis of the ESTs listed in Table 7.2. The relationship of these
15 contigs to the mouse cDNA sequence is illustrated in Figure 25.

Figure 28 is a diagrammatic representation of the relationship of sequence derived from analysis of mouse SOCS8 ESTs (Table 8.1 and Figure 29A) to the predicted protein structure of mouse SOCS8. The deduced partial amino acid sequence of mouse SOCS8 is shown in Figure 29B.
20 The structure of the protein is shown schematically with the SOCS box highlighted (). The predicted 3' untranslated region is shown by the thin line.

Figure 29A is a representation showing the partial nucleotide sequence of mouse SOCS8 cDNA (contig 8.1) derived from analysis of ESTs. The nucleotides encoding the part of the predicted
25 coding region, ending in the STOP codon are shown in upper case, while the predicted 3' untranslated regions are shown in lower case.

Figure 29B is a representation showing the partial predicted amino acid sequence of the mouse SOCS8 protein, derived from the nucleotide sequence in Figure 29A. The SOCS box, which
30 also shown in Figure 13 is underlined.

- 27 -

Figure 30 is a diagrammatic representation showing the relationship of mouse SOCS9 ESTs (Table 9.1) and human SOCS9 ESTs (Table 9.2). The nucleotide sequence of the mouse SOCS9 contig (m9.1) is shown in Figure 31, with the sequence of human SOCS9 contig (h9.1) being shown in Figure 32. The deduced amino acid sequence of human SOCS9 is shown
5 schematically, with the SH2 domain indicated by () and the SOCS box by (). The putative 3' untranslated region is shown by the thin solid line.

Figure 31 is a representation showing the partial nucleotide sequence of mouse SOCS9 cDNA (contig m9.1), derived from analysis of the ESTs listed in Table 9.1. The relationship of these
10 contigs to the mouse cDNA sequence is illustrated in Figure 30.

Figure 32 is a representation showing the partial nucleotide sequence of human SOCS9 cDNA (contig h9.1), derived from analysis of the ESTs listed in Table 9.2. Although it is clear that contig h9.1 encodes a protein with an SH2 domain and a SOCS box, the quality of the sequence
15 is not high enough to derive a single unambiguous open reading frame. The relationship of these contigs to the mouse cDNA sequence is illustrated in Figure 30.

Figure 33 is a representation showing the relationship of mouse SOCS10 cDNA clones (10-9, 10-12, 10-23 and 10-24) to contigs derived from analysis of mouse ESTs (Table 10.1) and
20 human ESTs (Table 10.2). The nucleotide sequence of the mouse SOCS10 contig is shown in Figure 10.2, with the sequence of human SOCS10 contigs (h10.1 and h10.2) being shown in Figure 35. The predicted structure of the protein is shown schematically, with the ankyrin repeats indicated by () and the SOCS box by (). The putative 3' untranslated regions is shown by the thin line solid line in the mouse and by the wavy line in h10.2. Based on analysis of clones
25 isolated to date and ESTs the 3' untranslated regions of mSOCS-10 and hSOCS-10 share little similarity.

Figure 34 is a representation showing the nucleotide sequence of the mouse SOCS10 derived from analysis of cDNA clone 10-9, 10-12, 10-23 and 10-24. The nucleotides encoding the part
30 of the predicted coding region, ending in the stop codon are shown in upper case, while the predicted 3' untranslated regions are shown in lower case. Although it is clear that contig m10.1

- 28 -

encodes a protein with a series of ankyrin repeats and a SOCS box, the quality of the sequence is not high enough to derive a single unambiguous open reading frame. The relationship of mouse cDNA sequence to mouse and human EST contigs is illustrated in Figure 33.

5 **Figure 35** is a representation showing the nucleotide sequence of human SOCS10 cDNA contig h10.2 and h10.2 derived from analysis of the ESTs listed in Table 10.2. The relationship of these contigs to the mouse cDNA sequence is illustrated in Figure 33.

Figure 36A is a representation showing the partial nucleotide sequence of the human SOCS11 cDNA derived from analysis of ESTs listed in Table 11.1 The nucleotides encoding the mature
10 coding region from the predicted ATG "start" codon to the stop codon is shown in upper case, while the predicted 5' and 3' untranslated regions are shown in lower case. The relationship of the partial cDNA sequence, derived from ESTs, to the predicted protein is shown in Figure 37.

Figure 36B is a representation showing the partial predicted amino acid sequence of human
15 SOCS11 protein, derived from the nucleotide sequence in Figure 36A. The SOCS box, which also shown in Figure 13, is underlined.

Figure 37 is a diagrammatic representation showing the relationship of sequence derived from analysis of human SOCS-11 ESTs (Table 11.1 and Figure 36A) to the predicted protein structure
20 of human SOCS11. The deduced partial amino acid sequence of human SOCS11 is shown in Figure 36B. The structure of the protein is shown schematically with the SH2 domain shown by () and the SOCS box highlighted by (). The predicted 3' untranslated region is shown by the thin line.

25 **Figure 38** is a diagrammatic representation showing the relationship of mouse SOCS12 cDNA clones (12-1) to contigs derived from analysis of mouse ESTs (Table 12.1) and human ESTs (Table 12.2). The nucleotide sequence of the mouse SOCS12 contig is shown in Figure 12.2, with the sequence of human SOCS12 contigs (h12.1 and h12.2) being shown in Figure 40. The deduced partial amino acid sequence of mouse SOCS12 is shown in Figure 39. The structure
30 of the protein is shown schematically, with the ankyrin repeats indicated by () and the SOCS box by (). The putative 3' untranslated region is shown by the thin line solid line in the mouse and

- 29 -

by the wavy line in h12.2. Based on analysis of clones isolated to date and ESTs the 3' untranslated regions of mSOCS12 and hSOCS12 share little similarity.

Figure 39 is a representation showing the nucleotide sequence of the mouse SOCS12 derived from analysis of cDNA clone 12-1 and the ESTs listed in Table 12.1. The nucleotides encoding the part of the predicted coding region, including the stop codon are shown in upper case, while the predicted 3' untranslated region is shown in lower case. By homology with human SOCS12 it is clear that contig m12.1 encodes a protein with a series of ankyrin repeats and a SOCS box, the quality of the sequence is not high enough to derive a single unambiguous open reading frame. The relationship of mouse cDNA sequence to mouse and human EST contigs is illustrated in Figure 38.

Figure 40 is a representation showing the nucleotide sequence of human SOCS12 cDNA contig h12.1 and h12.2 derived from analysis of the ESTs listed in Table 12.2. The relationship of these contigs to the mouse cDNA sequence is illustrated in Figure 38.

Figure 41 is a diagrammatic representation showing the relationship of contig m13.1 derived from analysis of mouse SOCS13 cDNA clones (62-1, 62-6-7, 62-14) and mouse ESTs (Table 13.1) to contig h13.1 derived from analysis of human ESTs (Table 13.2). The nucleotide sequence of the mouse SOCS13 contig is shown in Figure 42, with the sequence of human SOCS13 contig (h13.1) being shown in Figure 43. The deduced amino acid sequence of mouse SOCS13 is shown in Figure 42B. The structure of the protein is shown schematically, with the WD-40 repeats highlighted by () and the SOCS box highlighted by (). The 3' untranslated region is shown by the thin line solid line.

25

Figure 42A is a representation showing the nucleotide sequence of the mouse SOCS13 derived from analysis of cDNA clones 62-1, 62-6-7 and 62-14. The nucleotides encoding part of the predicted coding region, ending in the stop codon are shown in upper case, while those encoding the predicted 3' untranslated regions are shown in lower case. The relationship of mouse cDNA sequence to mouse and human EST contigs is illustrated in Figure 41.

30

- 30 -

Figure 42B is a representation showing the predicted amino acid sequence of mouse SOCS13 protein, derived from the nucleotide sequence in Figure 42A. The SOCS box, which also shown in Figure 13 is underlined.

5 **Figure 43** is a representation showing the nucleotide sequence of human SOCS13 cDNA contig h13.1 derived from analysis of the ESTs listed in Table 13.2. The relationship of these contigs to the mouse cDNA sequence is illustrated in Figure 41.

Figure 44 is a diagrammatic representation showing the relationship of a partial mouse SOCS14
10 cDNA clone (14-1) to contigs derived from analysis of mouse ESTs (Table 14.1). The nucleotide sequence of the mouse SOCS14 contig is shown in Figure 45. The deduced partial amino acid sequence of mouse SOCS14 is shown in Figure 45B. The structure of the protein is shown schematically, with the SH3 domain indicated by () and the SOCS box by (). The putative 3' untranslated region is shown by the thin line.

15

Figure 45A is a representation showing the nucleotide sequence of the mouse SOCS14 derived from analysis of genomic and cDNA clones. The nucleotides encoding the mature coding region from the predicted ATG "start" codon to the stop codon is shown in upper case, while the predicted 5' and 3' untranslated regions are shown in lower case. The relationship of mouse
20 cDNA sequence to mouse and human EST contigs is illustrated in Figure 44.

Figure 45B is a representation showing the predicted amino acid sequence of mouse SOCS14 protein, derived from the nucleotide sequence in Figure 45B. The SOCS box, which also shown in Figure 13 is underlined.

25

Figure 46 is a diagrammatic representation showing the relationship of contig m15.1 derived from analysis of mouse BAC and mouse ESTs (Table 15.1) to contig h15.1 derived from analysis of the human BAC and human ESTs (Table 15.2). The nucleotide sequence of the mouse SOCS15 contig is shown in Figure 47, with the sequence of human SOCS15 contig (h15.1)
30 being shown in Figure 47. The deduced amino acid sequence of mouse SOCS15 is shown in Figure 47B. The structure of the protein is shown schematically, with the WD-40 repeats

- 31 -

highlighted by () and the SOCS box highlighted by (). The 5' and 3' untranslated region are shown by the thin line solid line. The introns which interrupt the coding region are shown by ^.

Figure 47A is a representation showing the nucleotide sequence covering the mouse SOCS15 gene derived from analysis the mouse BAC listed in Table 15.1. The nucleotides encoding the predicted coding region, beginning with the ATG and ending in the stop codon are shown in upper case, while those encoding the predicted 5' untranslated region, the introns and the 3' untranslated region are shown in lower case. The relationship of mouse BAC to mouse and human ESTs contigs is illustrated in Figure 46.

10

Figure 47B is a representation showing the predicted amino acid sequence of mouse SOCS15 protein, derived from the nucleotide sequence in Figure 47A. The SOCS box, which also shown in Figure 13 is underlined.

Figure 48A is a representation showing the nucleotide sequence covering the human SOCS15 gene derived from analysis the human BAC listed in Table 15.2. The nucleotides encoding the predicted coding region, beginning with the ATG and ending in the stop codon are shown in upper case, while those encoding the predicted 5' untranslated region, the introns and the 3' untranslated region are shown in lower case. The relationship of the human BAC to mouse and human ESTs contigs is illustrated in Figure 46.

20

Figure 48B is a representation showing the predicted amino acid sequence of human SOCS15 protein, derived from the nucleotide sequence in Figure 48A. The SOCS box, which also shown in Figure 13 is underlined.

25

Figure 49 is a photographic representation showing SOCS1 inhibition of JAK2 kinase activity. (A) Upper panel. Cos M6 cells were transiently transfected with either Flag-tagged mJAK2 and mSOCS-1 DNA (SOCS1) or Flag-mJAK2 DNA alone (-), lysed, JAK2 proteins immunoprecipitated using anti-JAK2 antibody and subjected to an *in vitro* kinase assay. Lower panel. A portion of the JAK2 immunoprecipitates were Western blotted with anti-JAK2 antibody. (B) Upper panel. Cos M6 cells were transiently transfected with Flag- mJAK2 and

30

- 32 -

Flag- mSOCS-1 DNA or Flag-mJAK2 DNA alone, lysed, JAK2 proteins immunoprecipitated using anti-JAK2 (UBI) and separated by SDS/PAGE gel. Immunoprecipitates were then analysed by Western blot with anti-phosphotyrosine antibody. Lower panel; JAK2 expression. Cos cell lysates were separated by SDS/PAGE gel and analysed by Western blot with anti-FLAG
5 antibody (M2).

Figure 50 is a photographic representation showing interaction between JAK2 and SOCS protein. (A) Cos M6 cells were transiently transfected with Flag-tagged mJAK2 and various Flag-tagged SOCS DNAs (SOCS-1;S1, SOCS-2;S2, SOCS-3;S3, CIS) or Flag-mJAK2 alone,
10 lysed, JAK2 proteins immunoprecipitated using anti-JAK2 (UBI) and separated by SDS/PAGE. Immunoprecipitates were then analysed by Western blot with anti-FLAG antibody (M2). (B) Cos cell lysates described in (A) were separated by SDS/PAGE and expression levels of the various proteins were determined by Western blot with anti-FLAG antibody (M2). (C) JAK2 tyrosine phosphorylation. Cos cell lysates described in (A) were separated by SDS/PAGE and
15 proteins analysed by Western blot with anti-phosphotyrosine antibody.

Figure 51 is a diagrammatic representation of p β galpAloxneo.

Figure 52 is a diagrammatic representation of p β galpAloxneoTK.
20

Figure 53 is a diagrammatic representation of SOCS1 knockout construct.

25

30

- 33 -

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The present invention provides a new family of modulators of signal transduction. As the initial members of this family suppressed cytokine signalling, the family is referred to as the

5 "suppressors of cytokine signalling" family of "SOCS". The SOCS family is defined by the presence of a C-terminal domain referred to as a "SOCS box". Different classes of SOCS molecules are defined by a motif generally but not exclusively located N-terminal to the SOCS box and which is involved by protein:molecule interaction such as protein:DNA or protein:protein interaction. Particularly preferred motifs are selected from an SH2 domain, WD-

10 40 repeats and ankyrin repeats.

WD-40 repeats were originally recognised in the β -subunit of G-proteins. WD-40 repeats appear to form a β -propeller-like structure and may be involved in protein-protein interactions. Ankyrin repeats were originally recognised in the cytoskeletal protein ankryin.

15

Members of the SOCS family may be identified by any number of means. For example, SOCS1 to SOCS3 were identified by their ability to suppress cytokine-mediated signal transduction and, hence, were identified based on activity. SOCS4 to SOCS15 were identified as nucleotide sequences exhibiting similarity at the level of the SOCS box.

20

The SOCS box is a conserved motif located in the C-terminal region of the SOCS molecule. In accordance with the present invention, the amino acid sequence of the SOCS box is:

25

$$X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14} X_{15} X_{16} [X_i]_n X_{17} X_{18} X_{19} X_{20}$$

$$X_{21} X_{22} X_{23} [X_j]_n X_{24} X_{25} X_{26} X_{27} X_{28}$$

wherein:

- X_1 is L, I, V, M, A or P;
- X_2 is any amino acid residue;
- X_3 is P, T or S;
- X_4 is L, I, V, M, A or P;
- X_5 is any amino acid;

30

- 34 -

X_6 is any amino acid;
 X_7 is L, I, V, M, A, F, Y or W;
 X_8 is C, T or S;
 X_9 is R, K or H;
5 X_{10} is any amino acid;
 X_{11} is any amino acid;
 X_{12} is L, I, V, M, A or P;
 X_{13} is any amino acid;
 X_{14} is any amino acid;
10 X_{15} is any amino acid;
 X_{16} is L, I, V, M, A, P, G, C, T or S;
 $[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids
and wherein the sequence X_i may comprise the same or different amino
acids selected from any amino acid residue;
15 X_{17} is L, I, V, M, A or P;
 X_{18} is any amino acid;
 X_{19} is any amino acid;
 X_{20} L, I, V, M, A or P;
 X_{21} is P;
20 X_{22} is L, I, V, M, A, P or G;
 X_{23} is P or N;
 $[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids
and wherein the sequence X_j may comprise the same or different amino
acids selected from any amino acid residue;
25 X_{24} is L, I, V, M, A or P;
 X_{25} is any amino acid;
 X_{26} is any amino acid;
 X_{27} is Y or F; and
 X_{28} is L, I, V, M, A or P.

30

As stated above and in accordance with the present invention, SOCS proteins are divided into

- 35 -

separate classes based on the presence of a protein:molecule interacting region such as but not limited to an SH2 domain, WD-40 repeats and ankyrin repeats located N-terminal of the SOCS box. The latter three domains are protein:protein interacting domains.

- 5 Examples of SH2 containing SOCS proteins include SOCS1, SOCS2, SOCS3, SOCS5, SOCS9, SOCS11 and SOCS14. Examples of SOCS containing WD-40 repeats include SOCS4, SOCS6 and SOCS15. Examples of SOCS containing ankyrin repeats include SOCS7, SOCS10 and SOCS12.
- 10 The present invention provides *inter alia* nucleic acid molecules encoding SOCS proteins, purified naturally occurring SOCS proteins as well as recombinant forms of SOCS proteins and methods of modulating signal transduction by modulating activity of SOCS proteins or expression of SOCS genes. Preferably, signal transduction is mediated by a cytokine, examples of which include EPO, TPO, G-CSF, GM-CSF, IL-3, IL-2, IL-4, IL-7, IL-13, IL-6, LIF, IL-12,
- 15 IFN γ , TNF α , IL-1 and/or M-CSF. Particularly preferred cytokines include IL-6, LIF, OSM, IFN- γ and/or thrombopoietin.

Accordingly, one aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a

20 protein or a derivative, homologue, analogue or mimetic thereof or comprises a nucleotide sequence capable of hybridizing thereto under low stringency conditions at 42°C wherein said protein comprises a SOCS box in its C-terminal region and optionally a protein:molecule interacting domain N-terminal of the SOCS box.

- 25 Preferably, the protein:molecule interacting domain is a protein:DNA or protein:protein interacting domain. Most preferably, the protein:molecule interacting domain is one of an SH2 domain, WD-40 repeats and/or ankyrin repeats.

As stated above, preferably the subject SOCS modulate cytokine-mediated signal transduction.

30 The present invention extends, however, to SOCS molecules modulating other effector-mediated signal transduction such as mediated by other endogenous or exogenous molecules, antigens,

- 36 -

microbes and microbial products, viruses or components thereof, ions, hormones and parasites. Endogenous molecules in this context are molecules produced within the cell carrying the SOCS molecule. Exogenous molecules are produced by other cells or are introduced to the body.

- 5 Preferably, the nucleic acid molecule or SOCS protein is in isolated or purified form. The terms "isolated" and "purified" mean that a molecule has undergone at least one purification step away from other material.

Preferably, the nucleic acid molecule is in isolated form and is DNA such as cDNA or genomic
10 DNA. The DNA may encode the same amino acid sequence as the naturally occurring SOCS or the SOCS may contain one or more amino acid substitutions, deletions and/or additions. The nucleotide sequence may correspond to the genomic coding sequence (including exons and introns) or to the nucleotide sequence in cDNA from mRNA transcribed from the genomic gene or it may carry one or more nucleotide substitutions, deletions and/or additions thereto.

15

In a preferred embodiment, the nucleic acid molecule comprises a sequence of nucleotide encoding or complementary to a sequence encoding a SOCS protein or a derivative, homologue, analogue or mimetic thereof wherein the amino acid sequence of said SOCS protein is selected from SEQ ID NO:4 (mSOCS1), SEQ ID NO:6 (mSOCS2), SEQ ID NO:8 (mSOCS3), SEQ ID
20 NO:10 (hSOCS1), SEQ ID NO:12 (rSOCS1), SEQ ID NO:14 (mSOCS4), SEQ ID NO:18 (mSOCS5), SEQ ID NO:21 (mSOCS6), SEQ ID NO:25 (mSOCS27), SEQ ID NO:29 (mSOCS8), SEQ ID NO:36 (hSOCS11), SEQ ID NO:41 (mSOCS13), SEQ ID NO:44 (mSOCS14), SEQ ID NO:46 (mSOCS15) and SEQ ID NO:48 (mSOCS15) or encodes an amino acid sequence with a single or multiple amino acid substitution, deletion and/or addition to the
25 listed sequences or is a nucleotide sequence capable of hybridizing to the nucleic acid molecule under low stringency conditions at 42°C.

In an even more preferred embodiment, the present invention provides a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a
30 SOCS protein or a derivative, homologue, analogue or mimetic thereof wherein the nucleotide sequence is selected from a nucleotide sequence substantially set forth in SEQ ID NO:3

- 37 -

(mSOCS1), SEQ ID NO:5 (mSOCS2), SEQ ID NO:7 (mSOCS3), SEQ ID NO:9 (hSOCS11),
 SEQ ID NO:11 (rSOCS1), SEQ ID NO:13 (mSOCS4), SEQ ID NO:15 and SEQ ID NO:16
 (hSOCS4), SEQ ID NO:17 (mSOCS5), SEQ ID NO:19 (hSOCS5), SEQ ID NO:20 (mSOCS6),
 SEQ ID NO:22 and SEQ ID NO:23 (hSOCS6), SEQ ID NO:24 (mSOCS7), SEQ ID NO:26 and
 5 SEQ ID NO:27 (hSOCS7), SEQ ID NO:28 (mSOCS8), SEQ ID NO:30 (mSOCS9), SEQ ID
 NO:31 (hSOCS9), SEQ ID NO:32 (mSOCS10), SEQ ID NO:33 and SEQ ID NO:34
 (hSOCS10), SEQ ID NO:35 (hSOCS11), SEQ ID NO:37 (mSOCS12), SEQ ID NO:38 and
 SEQ ID NO:39 (hSOCS12), SEQ ID NO:40 (mSOCS13), SEQ ID NO:42 (hSOCS13), SEQ
 ID NO:43 (mSOCS14), SEQ ID NO:45 (mSOCS15) and SEQ ID NO:47 (hSOCS15) or a
 10 nucleotide sequence having at least about 15% similarity to all or a region of any of the listed
 sequences or a nucleic acid molecule capable of hybridizing to any of the listed sequences under
 low stringency conditions at 42°C.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1%
 15 v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for
 hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative
 stringency conditions may be applied where necessary, such as medium stringency, which
 includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and
 from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M
 20 to at least about 0.9M salt for washing conditions, or high stringency, which includes and
 encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least
 about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least
 about 0.15M salt for washing conditions.

25 In another embodiment, the present invention is directed to a SOCS protein or a derivative,
 homologue, analogue or mimetic thereof wherein said SOCS protein is identified as follows:

human SOCS4 characterised by EST81149, EST180909, EST182619, ya99H09,
 ye70co4, yh53c09, yh77g11, yh87h05, yi45h07, yj04e06, yq12h06, yq56a06, yq60e02,
 30 yq92g03, yq97h06, yr90f01, yt69c03, yv30a08, yv55f07, yv57h09, yv87h02, yv98e11,
 yw68d10, yw82a03, yx08a07, yx72h06, yx76b09, yy37h08, yy66b02, za81f08, zb18f07,

- 38 -

zc06e08, zd14g06, zd51h12, zd52b09, ze25g11, ze69f02, zf54f03, zh96e07, zv66h12,
zs83a08 and zs83g08;

mouse SOCS-4 characterised by mc65f04, mf42e06, mp10c10, mr81g09, and mt19h12;

5

human SOCS-5 characterised by EST15B103, EST15B105, EST27530 and zf50f01;

mouse SOCS-5 characterised by mc55a01, mh98f09, my26h12 and ve24e06;

10

human SOCS-6 characterised by yf61e08, yf93a09, yg05f12, yg41f04, yg45c02,
yh11f10, yh13b05, zc35a12, ze02h08, zl09a03, zl69e10, zn39d08 and zo39e06;

15

mouse SOCS-6 characterised by mc04c05, md48a03, mf31d03, mh26b07, mh78e11,
mh88h09, mh94h07, mi27h04 and mj29c05, mp66g04, mw75g03, va53b05, vb34h02,
vc55d07, vc59e05, vc67d03, vc68d10, vc97h01, vc99c08, vd07h03, vd08c01, vd09b12,
vd19b02, vd29a04 and vd46d06;

human SOCS-7 characterised by STS WI30171, EST00939, EST12913, yc29b05,
yp49f10, zt10f03 and zx73g04;

20

mouse SOCS-7 characterised by mj39a01 and vi52h07;

mouse SOCS-8 characterised by mj6e09 and vj27a029;

25

human SOCS-9 characterised by CSRL-82f2-u, EST114054, yy06b07, yy06g06,
zr40c09, zr72h01, yx92c08, yx93b08 and hfe0662;

mouse SOCS-9 characterised by me65d05;

30

human SOCS-10 characterised by aa48h10, zp35h01, zp97h12, zq08h01, zr34g05,
EST73000 and HSDHEI005;

- 39 -

mouse SOCS-10 characterised by mb14d12, mb40f06, mg89b11, mq89e12, mp03g12 and vh53c11;

human SOCS-11 characterised by zt24h06 and zr43b02;

5

human SOCS-13 characterised by EST59161;

mouse SOCS-13 characterised by ma39a09, me60c05, mi78g05, mk10c11, mo48g12, mp94a01, vb57c07 and vh07c11; and

10

human SOCS-14 characterised by mi75e03, vd29h11 and vd53g07;

or a derivative or homologue of the above ESTs characterised by a nucleic acid molecule being capable of hybridizing to any of the listed ESTs under low stringency conditions at 42°C.

15

In another embodiment, the nucleotide sequence encodes the following amino acid sequence:

$$X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14} X_{15} X_{16} [X_i]_n X_{17} X_{18} X_{19} X_{20} \\ X_{21} X_{22} X_{23} [X_j]_n X_{24} X_{25} X_{26} X_{27} X_{28}$$

20

wherein: X_1 is L, I, V, M, A or P;
 X_2 is any amino acid residue;
 X_3 is P, T or S;
 X_4 is L, I, V, M, A or P;
 X_5 is any amino acid;
 X_6 is any amino acid;
 X_7 is L, I, V, M, A, F, Y or W;
 X_8 is C, T or S;
 X_9 is R, K or H;
 X_{10} is any amino acid;
 X_{11} is any amino acid;

25

30

- 40 -

X_{12} is L, I, V, M, A or P;

X_{13} is any amino acid;

X_{14} is any amino acid;

X_{15} is any amino acid;

5 X_{16} is L, I, V, M, A, P, G, C, T or S;

$[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;

X_{17} is L, I, V, M, A or P;

10 X_{18} is any amino acid;

X_{19} is any amino acid;

X_{20} L, I, V, M, A or P;

X_{21} is P;

X_{22} is L, I, V, M, A, P or G;

15 X_{23} is P or N;

$[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_j may comprise the same or different amino acids selected from any amino acid residue;

X_{24} is L, I, V, M, A or P;

20 X_{25} is any amino acid;

X_{26} is any amino acid;

X_{27} is Y or F; and

X_{28} is L, I, V, M, A or P.

25 The above sequence comparisons are preferably to the whole molecule but may also be to part thereof. Preferably, the comparisons are made to a contiguous series of at least about 21 nucleotides or at least about 5 amino acids. More preferably, the comparisons are made against at least about 21 contiguous nucleotides or at least 7 contiguous amino acids. Comparisons may also only be made to the SOCS box region or a region encompassing the protein:molecule
30 interacting region such as the SH2 domain WD-40 repeats and/or ankyrin repeats.

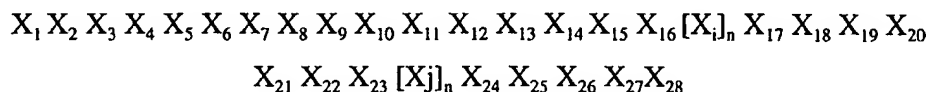
- 41 -

Still another embodiment of the present invention contemplates an isolated polypeptide or a derivative, homologue, analogue or mimetic thereof comprising a SOCS box in its C-terminal region.

- 5 Preferably the polypeptide further comprises a protein:molecule interacting domain such as a protein:DNA or protein:protein interacting domain. Preferably, this domain is located N-terminal of the SOCS box. It is particularly preferred for the protein:molecule interacting domain to be at least one of an SH2 domain, WD-40 repeats and/or ankyrin repeats.
- 10 Preferably, the signal transduction is mediated by a cytokine selected from EPO, TPO, G-CSF, GM-CSF, IL-3, IL-2, IL-4, IL-7, IL-13, IL-6, LIF, IL-12, IFN γ , TNF α , IL-1 and/or M-CSF. Preferred cytokines are IL-6, LIF, OSM, IFN- γ or thrombopoietin.

More preferably, the protein comprises a SOCS box having the amino acid sequence:

15



wherein:

20 X_1 is L, I, V, M, A or P;
 X_2 is any amino acid residue;
 X_3 is P, T or S;
 X_4 is L, I, V, M, A or P;
 X_5 is any amino acid;
 X_6 is any amino acid;
25 X_7 is L, I, V, M, A, F, Y or W;
 X_8 is C, T or S;
 X_9 is R, K or H;
 X_{10} is any amino acid;
 X_{11} is any amino acid;
30 X_{12} is L, I, V, M, A or P;
 X_{13} is any amino acid;

- 42 -

- X_{14} is any amino acid;
 X_{15} is any amino acid;
 X_{16} is L, I, V, M, A, P, G, C, T or S;
 $[X_i]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids
 5 and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;
 X_{17} is L, I, V, M, A or P;
 X_{18} is any amino acid;
 X_{19} is any amino acid;
 10 X_{20} L, I, V, M, A or P;
 X_{21} is P;
 X_{22} is L, I, V, M, A, P or G;
 X_{23} is P or N;
 $[X_i]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids
 15 and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;
 X_{24} is L, I, V, M, A or P;
 X_{25} is any amino acid;
 X_{26} is any amino acid;
 20 X_{27} is Y or F; and
 X_{28} is L, I, V, M, A or P.

Still another embodiment provides an isolated polypeptide or a derivative, homologue, analogue or mimetic thereof comprising a sequence of amino acids substantially as set forth in SEQ ID
 25 NO:4 (mSOCS1), SEQ ID NO:6 (mSOCS2), SEQ ID NO:8 (mSOCS3), SEQ ID NO:10 (hSOCS1), SEQ ID NO:12 (rSOCS1), SEQ ID NO:14 (mSOCS4), SEQ ID NO:18 (mSOCS5), SEQ ID NO:21 (mSOCS6), SEQ ID NO:25 (mSOCS7), SEQ ID NO:29 (mSOCS8), SEQ ID NO:36 (hSOCS11), SEQ ID NO:41 (mSOCS13), SEQ ID NO:44 (mSOCS14), SEQ ID NO:46 (mSOCS15) and SEQ ID NO:48 (hSOCS15) or an amino acid sequence having at least 15%
 30 similarity to all or a part of the listed sequences.

- 43 -

Preferred nucleotide percentage similarities include at least about 20%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90% or above such as 93%, 95%, 98% or 99%.

- 5 Preferred amino acid similarities include at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 97% or 98% or above.

As stated above, similarity may be measured against an entire molecule or a region comprising
10 at least 21 nucleotides or at least 7 amino acids. Preferably, similarity is measured in a conserved region such as SH2 domain, WD-40 repeats, ankyrin repeats or other protein:molecule interacting domains or a SOCS box.

The term "similarity" includes exact identity between sequences or, where the sequence differs,
15 different amino acids are related to each other at the structural, functional, biochemical and/or conformational levels.

The nucleic acid molecule may be isolated from any animal such as humans, primates, livestock animals (e.g. horses, cows, sheep, donkeys, pigs), laboratory test animals (e.g. mice, rats, rabbits,
20 hamsters, guinea pigs), companion animals (e.g. dogs, cats) or captive wild animals (e.g. deer, foxes, kangaroos).

The terms "derivatives" or its singular form "derivative" whether in relation to a nucleic acid molecule or a protein includes parts, mutants, fragments and analogues as well as hybrid or
25 fusion molecules and glycosylation variants. Particularly useful derivatives comprise single or multiple amino acid substitutions, deletions and/or additions to the SOCS amino acid sequence.

Preferably, the derivatives have functional activity or alternatively act as antagonists or agonists. The present invention further extends to homologues of SOCS which include the functionally or
30 structurally related molecule from different animal species. The present invention also encompasses analogues and mimetics. Mimetics include a class of molecule generally but not

- 44 -

necessarily having a non-amino acid structure and which functionally are capable of acting in an analogous manner to the protein for which it is a mimic, in this case, a SOCS. Mimetics may comprise a carbohydrate, aromatic ring, lipid or other complex chemical structure or may also be proteinaceous in composition. Mimetics as well as agonists and antagonists contemplated
5 herein are conveniently located through systematic searching of environments, such as coral, marine and freshwater river beds, flora and microorganisms. This is sometimes referred to as natural product screening. Alternatively, libraries of synthetic chemical compounds may be screened for potentially useful molecules.

10 As stated above, the present invention contemplates agonists and antagonists of the SOCS. One example of an antagonist is an antisense oligonucleotide sequence. Useful oligonucleotides are those which have a nucleotide sequence complementary to at least a portion of the protein-coding or "sense" sequence of the nucleotide sequence. These anti-sense nucleotides can be used to effect the specific inhibition of gene expression. The antisense approach can cause
15 inhibition of gene expression apparently by forming an anti-parallel duplex by complementary base pairing between the antisense construct and the targeted mRNA, presumably resulting in hybridisation arrest of translation. Ribozymes and co-suppression molecules may also be used. Antisense and other nucleic acid molecules may first need to be chemically modified to permit penetration of cell membranes and/or to increase their serum half life or otherwise make them
20 more stable for *in vivo* administration. Antibodies may also act as either antagonists or agonists although are more useful in diagnostic applications or in the purification of SOCS proteins. Antagonists and agonists may also be identified following natural product screening or screening of libraries of chemical compounds or may be derivatives or analogues of the SOCS molecules.

25

Accordingly, the present invention extends to analogues of the SOCS proteins of the present invention. Analogues may be used, for example, in the treatment or prophylaxis of cytokine mediated dysfunction such as autoimmunity, immune suppression or hyperactive immunity or other condition including but not limited to dysfunctions in the haemopoietic, endocrine, hepatic
30 and neural systems. Dysfunctions mediated by other signal transducing elements such as hormones or endogenous or exogenous molecules, antigens, microbes and microbial products,

- 45 -

viruses or components thereof, ions, hormones and parasites are also contemplated by the present invention.

Analogues of the proteins contemplated herein include, but are not limited to, modification to
5 side chains, incorporating of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogues.

Examples of side chain modifications contemplated by the present invention include
10 modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with NaBH_4 ; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-
15 phosphate followed by reduction with NaBH_4 .

The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

20 The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitisation, for example, to a corresponding amide.

Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides
25 with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline pH.

30 Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides.

- 46 -

Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with
5 iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine,
10 sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acid, contemplated herein is shown in Table 3.

- 47 -

TABLE 3

Non-conventional amino acid		Non-conventional amino acid	
	Code		Code
5			
α -aminobutyric acid	Abu	L-N-methylalanine	Nmala
α -amino- α -methylbutyrate	Mgab	L-N-methylarginine	Nmarg
aminocyclopropane-	Cpro	L-N-methylasparagine	Nmasn
10 carboxylate		L-N-methylaspartic acid	Nmasp
aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
aminonorbornyl-	Norb	L-N-methylglutamine	Nmgln
carboxylate		L-N-methylglutamic acid	Nmglu
cyclohexylalanine		Chexa L-N-methylhistidine	Nmhis
15 cyclopentylalanine	Cpen	L-N-methylisoleucine	Nmile
D-alanine	Dal	L-N-methylleucine	Nmleu
D-arginine	Darg	L-N-methyllysine	Nmlys
D-aspartic acid	Das	L-N-methylmethionine	Nmmet
D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
20 D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
D-isoleucine	Dile	L-N-methylproline	Nmpro
D-leucine	Dleu	L-N-methylserine	Nmser
25 D-lysine	Dlys	L-N-methylthreonine	Nmthr
D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
D-phenylalanine	Dphe	L-N-methylvaline	Nmval
D-proline	Dpro	L-N-methylethylglycine	Nmetg
30 D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
D-threonine	Dthr	L-norleucine	Nle

- 48 -

	D-tryptophan	Dtrp	L-norvaline	Nva
	D-tyrosine	Dtyr	α -methyl-aminoisobutyrate	Maib
	D-valine	Dval	α -methyl- γ -aminobutyrate	Mgabu
	D- α -methylalanine	Dmala	α -methylcyclohexylalanine	Mchexa
5	D- α -methylarginine	Dmarg	α -methylcyclopentylalanine	Mcpen
	D- α -methylasparagine	Dmasn	α -methyl- α -naphthylalanine	Manap
	D- α -methylaspartate	Dmasp	α -methylpenicillamine	Mpen
	D- α -methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D- α -methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
10	D- α -methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
	D- α -methylisoleucine	Dmile	N-amino- α -methylbutyrate	Nmaabu
	D- α -methyllleucine	Dmleu	α -naphthylalanine	Anap
	D- α -methyllysine	Dmlys	N-benzylglycine	Nphe
	D- α -methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
15	D- α -methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
	D- α -methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D- α -methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D- α -methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D- α -methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
20	D- α -methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
	D- α -methyltyrosine	Dmtty	N-cyclodecylglycine	Ncdec
	D- α -methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
25	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
30	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl)glycine	Nser
	D-N-methylisoleucine	Dnmile	N-(imidazolylethyl)glycine	Nhis

- 49 -

	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp
	D-N-methyllysine	Dnmlys	N-methyl- γ -aminobutyrate	Nmgabu
	N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmt
	D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
5	N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
	N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
	N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
	N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
	D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
10	D-N-methyltyrosine	Dnmtyr	N-methyl- α -naphthylalanine	Nmanap
	D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
	γ -aminobutyric acid	Gabu	N-(<i>p</i> -hydroxyphenyl)glycine	Nhtyr
	L- <i>t</i> -butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
	L-ethylglycine	Etg	penicillamine	Pen
15	L-homophenylalanine	Hphe	L- α -methylalanine	Mala
	L- α -methylarginine	Marg	L- α -methylasparagine	Masn
	L- α -methylaspartate	Masp	L- α -methyl- <i>t</i> -butylglycine	Mtbug
	L- α -methylcysteine	Mcys	L-methylethylglycine	Metg
	L- α -methylglutamine	Mgln	L- α -methylglutamate	Mglu
20	L- α -methylhistidine	Mhis	L- α -methylhomophenylalanine	Mhphe
	L- α -methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
	L- α -methylleucine	Mleu	L- α -methyllysine	Mlys
	L- α -methylmethionine	Mmet	L- α -methylnorleucine	Mnle
	L- α -methylnorvaline	Mnva	L- α -methylornithine	Morn
25	L- α -methylphenylalanine	Mphe	L- α -methylproline	Mpro
	L- α -methylserine	Mser	L- α -methylthreonine	Mthr
	L- α -methyltryptophan	Mtrp	L- α -methyltyrosine	Mtyr
	L- α -methylvaline	Mval	L-N-methylhomophenylalanine	Nmhph

- 50 -

N-(N-(2,2-diphenylethyl)	Nnbhm	N-(N-(3,3-diphenylpropyl)	Nnbhe
carbamylmethyl)glycine		carbamylmethyl)glycine	
1-carboxy-1-(2,2-diphenyl-	Nmbc		
ethylamino)cyclopropane			

5

Crosslinkers can be used, for example, to stabilise 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with $n=1$ to $n=6$, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually

10 contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of C_α and N_α -methylamino acids, introduction of double bonds between C_α and C_β atoms of amino acids and the formation of cyclic peptides or analogues by introducing covalent bonds such as forming

15 an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

20

These types of modifications may be important to stabilise the cytokines if administered to an individual or for use as a diagnostic reagent.

25

Other derivatives contemplated by the present invention include a range of glycosylation variants from a completely unglycosylated molecule to a modified glycosylated molecule. Altered glycosylation patterns may result from expression of recombinant molecules in different host cells.

30

Another embodiment of the present invention contemplates a method for modulating expression of a SOCS protein in a mammal, said method comprising contacting a gene encoding a SOCS or a factor/element involved in controlling expression of the SOCS gene with an effective amount of a modulator of SOCS expression for a time and under conditions sufficient

to up-regulate or down-regulate or otherwise modulate expression of SOCS. An example of a modulator is a cytokine such as IL-6 or other transcription regulators of SOCS expression.

- 51 -

Expression includes transcription or translation or both.

Another aspect of the present invention contemplates a method of modulating activity of SOCS in a human, said method comprising administering to said mammal a modulating effective
5 amount of a molecule for a time and under conditions sufficient to increase or decrease SOCS activity. The molecule may be a proteinaceous molecule or a chemical entity and may also be a derivative of SOCS or a chemical analogue or truncation mutant of SOCS.

A further aspect of the present invention provides a method of inducing synthesis of a SOCS
10 or transcription/translation of a SOCS comprising contacting a cell containing a SOCS gene with an effective amount of a cytokine capable of inducing said SOCS for a time and under conditions sufficient for said SOCS to be produced. For example, SOCS1 may be induced by IL-6.

15 Still a further aspect of the present invention contemplates a method of modulating levels of a SOCS protein in a cell said method comprising contacting a cell containing a SOCS gene with an effective amount of a modulator of SOCS gene expression or SOCS protein activity for a time and under conditions sufficient to modulate levels of said SOCS protein.

20 Yet a further aspect of the present invention contemplates a method of modulating signal transduction in a cell containing a SOCS gene comprising contacting said cell with an effective amount of a modulator of SOCS gene expression or SOCS protein activity for a time sufficient to modulate signal transduction.

25 Even yet a further aspect of the present invention contemplates a method of influencing interaction between cells wherein at least one cell carries a SOCS gene, said method comprising contacting the cell carrying the SOCS gene with an effective amount of a modulator of SOCS gene expression or SOCS protein activity for a time sufficient to modulate signal transduction.

30 As stated above, of the present invention contemplates a range of mimetics or small molecules capable of acting as agonists or antagonists of the SOCS. Such molecules may be obtained

- 52 -

from natural product screening such as from coral, soil, plants or the ocean or antarctic environments. Alternatively, peptide, polypeptide or protein libraries or chemical libraries may be readily screened. For example, M1 cells expressing a SOCS do not undergo differentiation in the presence of IL-6. This system can be used to screen molecules which permit
5 differentiation in the presence of IL-6 and a SOCS. A range of test cells may be prepared to screen for antagonists and agonists for a range of cytokines. Such molecules are preferably small molecules and may be of amino acid origin or of chemical origin. SOCS molecules interacting with signalling proteins (eg. JAKS) provide molecular screens to detect molecules which interfere or promote this interaction. Once such screening protocol involves natural
10 product screening.

Accordingly, the present invention contemplates a pharmaceutical composition comprising SOCS or a derivative thereof or a modulator of SOCS expression or SOCS activity and one or more pharmaceutically acceptable carriers and/or diluents. These components are referred to
15 as the "active ingredients". These and other aspects of the present invention apply to any SOCS molecules such as but not limited to SOCS1 to SOCS15.

The pharmaceutical forms containing active ingredients suitable for injectable use include sterile aqueous solutions (where water soluble) sterile powders for the extemporaneous preparation
20 of sterile injectable solutions. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for
25 example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride.
30 Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

- 53 -

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions is such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.1 μ g and 2000 mg of active compound.

20

The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter. A binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such as sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any

30

- 54 -

dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound(s) may be incorporated into sustained-release preparations and formulations.

- 5 The present invention also extends to forms suitable for topical application such as creams, lotions and gels.

Pharmaceutically acceptable carriers and/or diluents include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and
10 the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

- 15 It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for
20 the novel dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active material for the treatment of disease in living subjects having a diseased condition in which bodily health is impaired as herein disclosed in detail.

25

The principal active ingredient is compounded for convenient and effective administration in effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form as hereinbefore disclosed. A unit dosage form can, for example, contain the principal active compound in amounts ranging from 0.5 μg to about 2000 mg. Expressed in proportions, the
30 active compound is generally present in from about 0.5 μg to about 2000 mg/ml of carrier. In the case of compositions containing supplementary active ingredients, the dosages are

- 55 -

determined by reference to the usual dose and manner of administration of the said ingredients. The effective amount may also be conveniently expressed in terms of an amount per kg of body weight. For example, from about 0.01 ng to about 10,000 mg/kg body weight may be administered.

5

The pharmaceutical composition may also comprise genetic molecules such as a vector capable of transfecting target cells where the vector carries a nucleic acid molecule capable of modulating SOCS expression or SOCS activity. The vector may, for example, be a viral vector. In this regard, a range of gene therapies are contemplated by the present invention including
10 isolating certain cells, genetically manipulating and returning the cell to the same subject or to a genetically related or similar subject.

Still another aspect of the present invention is directed to antibodies to SOCS and its derivatives. Such antibodies may be monoclonal or polyclonal and may be selected from
15 naturally occurring antibodies to SOCS or may be specifically raised to SOCS or derivatives thereof. In the case of the latter, SOCS or its derivatives may first need to be associated with a carrier molecule. The antibodies and/or recombinant SOCS or its derivatives of the present invention are particularly useful as therapeutic or diagnostic agents.

20 For example, SOCS and its derivatives can be used to screen for naturally occurring antibodies to SOCS. These may occur, for example in some autoimmune diseases. Alternatively, specific antibodies can be used to screen for SOCS. Techniques for such assays are well known in the art and include, for example, sandwich assays and ELISA. Knowledge of SOCS levels may be important for diagnosis of certain cancers or a predisposition to cancers or monitoring cytokine
25 mediated cellular responsiveness or for monitoring certain therapeutic protocols.

Antibodies to SOCS of the present invention may be monoclonal or polyclonal. Alternatively, fragments of antibodies may be used such as Fab fragments. Furthermore, the present invention extends to recombinant and synthetic antibodies and to antibody hybrids. A "synthetic
30 antibody" is considered herein to include fragments and hybrids of antibodies. The antibodies of this aspect of the present invention are particularly useful for immunotherapy and may also

- 56 -

be used as a diagnostic tool for assessing apoptosis or monitoring the program of a therapeutic regimen.

For example, specific antibodies can be used to screen for SOCS proteins. The latter would be
5 important, for example, as a means for screening for levels of SOCS in a cell extract or other biological fluid or purifying SOCS made by recombinant means from culture supernatant fluid. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

10 It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of SOCS.

15

Both polyclonal and monoclonal antibodies are obtainable by immunization with the enzyme or protein and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable laboratory animal with an effective amount of SOCS, or
20 antigenic parts thereof, collecting serum from the animal, and isolating specific sera by any of the known immunoabsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

25 The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques which are well known to those who are skilled in the art.

30

Another aspect of the present invention contemplates a method for detecting SOCS in a

- 57 -

biological sample from a subject said method comprising contacting said biological sample with an antibody specific for SOCS or its derivatives or homologues for a time and under conditions sufficient for an antibody-SOCS complex to form and then detecting said complex.

- 5 The presence of SOCS may be accomplished in a number of ways such as by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to US Patent Nos. 4,016,043, 4,424,279 and 4,018,653. These, of course, include both single-site and two-site or "sandwich" assays of the non-competitive types, as well as in the traditional competitive binding assays. These assays also include direct binding of a labelled
10 antibody to a target.

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay,
15 an unlabelled antibody is immobilized on a solid substrate and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-
20 antigen-labelled antibody. Any unreacted material is washed away, and the presence of the antigen is determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are
25 added simultaneously to the bound antibody. These techniques are well known to those skilled in the art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain SOCS including cell extract, tissue biopsy or possibly serum, saliva, mucosal secretions, lymph, tissue fluid and respiratory fluid. The sample is, therefore, generally a biological sample comprising biological fluid but also
30 extends to fermentation fluid and supernatant fluid such as from a cell culture.

- 58 -

In the typical forward sandwich assay, a first antibody having specificity for the SOCS or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports
5 may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40
10 minutes or overnight if more convenient) and under suitable conditions (e.g. room temperature to 37°C) to allow binding of any subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

15

An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody.
20 Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its
25 chemical nature, provides an analytically identifiable signal which allows the detection of antigen-bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

30 In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a

- 59 -

wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding
5 enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then
10 added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

15

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a
20 characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescence and EIA techniques are both very well established in the art and are
25 particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

The present invention also contemplates genetic assays such as involving PCR analysis to detect SOCS gene or its derivatives. Alternative methods or methods used in conjunction include
30 direct nucleotide sequencing or mutation scanning such as single stranded conformation polymorphisms analysis (SSCP) as specific oligonucleotide hybridisation, as methods such as

- 60 -

direct protein truncation tests.

Since cytokines are involved in transcription of some SOCS molecules, the detection of SOCS provides surrogate markers for cytokines or cytokine activity. This may be useful in assessing
5 subjects with a range of conditions such as those will autoimmune diseases, for example, rheumatoid arthritis, diabetes and stiff man syndrome amongst others.

The nucleic acid molecules of the present invention may be DNA or RNA. When the nucleic acid molecule is in DNA form, it may be genomic DNA or cDNA. RNA forms of the nucleic
10 acid molecules of the present invention are generally mRNA.

Although the nucleic acid molecules of the present invention are generally in isolated form, they may be integrated into or ligated to or otherwise fused or associated with other genetic molecules such as vector molecules and in particular expression vector molecules. Vectors and
15 expression vectors are generally capable of replication and, if applicable, expression in one or both of a prokaryotic cell or a eukaryotic cell. Preferably, prokaryotic cells include *E. coli*, *Bacillus sp* and *Pseudomonas sp*. Preferred eukaryotic cells include yeast, fungal, mammalian and insect cells.

20 Accordingly, another aspect of the present invention contemplates a genetic construct comprising a vector portion and a mammalian and more particularly a human SOCS gene portion, which SOCS gene portion is capable of encoding a SOCS polypeptide or a functional or immunologically interactive derivative thereof.

25 Preferably, the SOCS gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said SOCS gene portion in an appropriate cell.

In addition, the SOCS gene portion of the genetic construct may comprise all or part of the
30 gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

- 61 -

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

The present invention also extends to any or all derivatives of SOCS including mutants, part,
5 fragments, portions, homologues and analogues or their encoding genetic sequence including single or multiple nucleotide or amino acid substitutions, additions and/or deletions to the naturally occurring nucleotide or amino acid sequence. The present invention also extends to mimetics and agonists and antagonists of SOCS.

10 The SOCS and its genetic sequence of the present invention will be useful in the generation of a range of therapeutic and diagnostic reagents and will be especially useful in the detection of a cytokine involved in a particular cellular response or a receptor for that cytokine. For example, cells expressing SOCS gene such as M1 cells expressing the SOCS1 gene, will no longer be responsive to a particular cytokine such as, in the case of SOCS1, IL-6. Clearly, the
15 present invention further contemplates cells such as M1 cells expressing any SOCS gene such as from SOCS1 to SOCS15. Furthermore, the present invention provides the use of molecules that regulate or potentiate the ability of therapeutic cytokines. For example, molecules which block some SOCS activity, may act to potential therapeutic cytokine activity (eg. G-CSF).

20 Soluble SOCS polypeptides are also contemplated to be particularly useful in the treatment of disease, injury or abnormality involving cytokine mediated cellular responsiveness such as hyperimmunity, immunosuppression, allergies, hypertension and the like.

A further aspect of the present invention contemplates the use of SOCS or its functional
25 derivatives in the manufacture of a medicament for the treatment of conditions involving cytokine mediated cellular responsiveness.

The present invention further contemplates transgenic mammalian cells expressing a SOCS gene. Such cells are useful indicator cell lines for assaying for suppression of cytokine function.
30 One example is M1 cells expressing a SOCS gene. Such cell lines may be useful for screening for cytokines or screening molecules such as naturally occurring molecules from plants, coral,

- 62 -

microorganisms or bio-organically active soil or water capable of acting as cytokine antagonists or agonists.

The present invention further contemplates hybrids between different SOCS from the same or
5 different animal species. For example, a hybrid may be formed between all or a functional part of mouse SOCS1 and human SOCS1. Alternatively, the hybrid may be between all or part of mouse SOCS1 and mouse SOCS2. All such hybrids are contemplated herein and are particularly useful in developing pleiotropic molecules.

10 The present invention further contemplates a range of genetic based diagnostic assays screening for individuals with defective SOCS genes. Such mutations may result in cell types not being responsive to a particular cytokine or resulting in over responsiveness leading to a range of conditions. The SOCS genetic sequence can be readily verified using a range of PCR or other techniques to determine whether a mutation is resident in the gene. Appropriate gene therapy
15 or other interventionist therapy may then be adopted.

The present invention is further described by the following non-limiting Examples.

- 63 -

Examples 1-16 relate to SOCS1, SOCS2 and SOCS3 which were identified on the basis of activity. Examples 17-24 relate to various aspects of SOCS4 to SOCS15 which were cloned initially on the basis of sequence similarity. Examples 25-36 relate to specific aspects of SOCS4 to SOCS15, respectively.

5

EXAMPLE 1**CELL CULTURE AND CYTOKINES**

The M1 cell line was derived from a spontaneously arising leukaemia in SL mice [Ichikawa, 1969]. Parental M1 cells used in this study have been in passage at the Walter and Eliza Hall Institute for Medical Research, Melbourne, Victoria, Australia, for approximately 10 years. M1
10 cells were maintained by weekly passage in Dulbecco's modified Eagle's medium (DME) containing 10% (v/v) foetal bovine serum (FCS). Recombinant cytokines are generally available from commercial sources or were prepared by published methods. Recombinant murine LIF was produced in *Escherichia coli* and purified, as previously described [Gearing, 1989]. Purified human oncostatin M was purchased from PeproTech Inc (Rocky Hill, NJ,
15 USA), and purified mouse IFN- γ was obtained from Genzyme Diagnostics (Cambridge, MA, USA). Recombinant murine thrombopoietin was produced as a FLAGTM-tagged fusion protein in CHO cells and then purified.

20

EXAMPLE 2**AGAR COLONY ASSAYS**

In order to assay the differentiation of M1 cells in response to cytokines, 300 cells were cultured in 35 mm Petri dishes containing 1 ml of DME supplemented with 20%(v/v) fetal calf serum (FCS), 0.3%(w/v) agar and 0.1 ml of serial dilutions of IL-6, LIF, OSM, IFN- γ , tpo or dexamethasone (Sigma Chemical Company, St Louis, MI). After 7 days culture at 37°C in a
25 fully humidified atmosphere, containing 10% (v/v) CO₂ in air, colonies of M1 cells were counted and classified as differentiated if they were composed of dispersed cells or had a corona of dispersed cells around a tightly packed centre.

30

EXAMPLE 3**GENERATION OF RETROVIRAL LIBRARY**

A cDNA expression library was constructed from the factor-dependent haemopoietic cell line

- 64 -

FDC-P1, essentially as described [Rayner, 1994]. Briefly, cDNA was cloned into the retroviral vector pRUFneo and then transfected into an amphotrophic packaging cell line (PA317). Transiently generated virus was harvested from the cell supernatant at 48 hr posttransfection, and used to infect Y2 ecotropic packaging cells, to generate a high titre virus-producing cell line.

EXAMPLE 4

RETROVIRAL INFECTION OF M1 CELLS

Pools of 10^6 infected Ψ 2 cells were irradiated (3000 rad) and cocultivated with 10^6 M1 cells in DME supplemented with 10%(v/v) FCS and 4 μ g/ml Polybrene, for 2 days at 37°C. To select for IL-6-unresponsive clones, retrovirally-infected M1 cells were washed once in DME, and cultured at approximately 2×10^4 cells/ml in 1 ml agar cultures containing 400 μ g/ml geneticin (GibcoBRL, Grand Island, NY) and 100 ng/ml IL-6. The efficiency of infection of M1 cells was 1-2%, as estimated by agar plating the infected cells in the presence of geneticin only.

EXAMPLE 5

PCR

Genomic DNA from retrovirally-infected M1 cells was digested with Sac I and 1 μ g of phenol/chloroform extracted DNA was then amplified by polymerase chain reaction (PCR). Primers used for amplification of cDNA inserts from the integrated retrovirus were GAG3 (5' CACGCCGCCCCACGTGAAGGC 3' [SEQ ID NO:1]), which corresponds to the vector gag sequence approximately 30 bp 5' of the multiple cloning site, and HSVTK (5' TTCGCCAATGACAAGACGCT 3' [SEQ ID NO:2]), which corresponds to the pMC1neo sequence approximately 200 bp 3' of the multiple cloning site. The PCR entailed an initial denaturation at 94°C for 5 min, 35 cycles of denaturation at 94°C for 1 min, annealing at 56°C for 2 min, and extension at 72°C for 3 min, followed by a final 10 min extension. PCR products were gel purified and then ligated into the pGEM-T plasmid (Promega, Madison, WI), and sequenced using an ABI PRISM Dye Terminator Cycle Sequencing Kit and a Model 373 Automated DNA Sequencer (Applied Biosystems Inc., Foster City, CA).

- 65 -

EXAMPLE 6**CLONING OF cDNAs**

Independent cDNA clones encoding mouse SOCS1 were isolated from a murine thymus cDNA library essentially as described (Hilton *et al*, 1994). The nucleotide and predicted amino acid sequences of mouse SOCS1 cDNA were compared to databases using the BLASTN and TFASTA algorithms (Pearson and Lipman, 1988; Pearson, 1990; Altshul *et al*, 1990). Oligonucleotides were designed from the ESTs encoding human SOCS1 and mouse SOC-1 and SOCS3 and used to probe commercially available mouse thymus and spleen cDNA libraries. Sequencing was performed using an ABI automated sequencer according to the manufacturer's instructions.

EXAMPLE 7**SOUTHERN AND NORTHERN BLOT ANALYSES AND RT-PCR**

³²P-labelled probes were generated using a random decanucleotide labelling kit (Bresatec, Adelaide, South Australia) from a 600 bp Pst I fragment encoding neomycin phosphotransferase from the plasmid pPGKneo, 1070 bp fragment of the SOCS1 gene obtained by digestion of the 1.4 kbp PCR product with Xho I, SOCS2, SOCS3, CIS and a 1.2 kbp fragment of the chicken glyceraldehyde 3-phosphate dehydrogenase gene [Dugaiczky, 1983].

Genomic DNA was isolated from cells using a proteinase K-sodium dodecyl sulfate procedure essentially as described. Fifteen micrograms of DNA was digested with either BamH I or Sac I, fractionated on a 0.8%(w/v) agarose gel, transferred to GeneScreenPlus membrane (Du Pont NEN, Boston MA), prehybridised, hybridised with random-primed ³²P-labelled DNA fragments and washed essentially as described [Sambrook, 1989].

25

Total RNA was isolated from cells and tissues using Trizol Reagent, as recommended by the manufacturer (GibcoBRL, Grand Island, NY). When required polyA+ mRNA was purified essentially as described [Alexander, 1995]. Northern blots were prehybridised, hybridized with random-primed ³²P-labelled DNA fragments and washed as described [Alexander, 1995].

30

To assess the induction of SOCS genes by IL-6, mice (C57BL6) were injected intravenously

- 66 -

with 5 μ g IL-6 followed by harvest of the liver at the indicated timepoints after injection. M1 cells were cultured in the presence of 20 ng/ml IL-6 and harvested at the indicated times. For RT-PCR analysis, bone marrow cells were harvested as described (Metcalf *et al*, 1995) and stimulated for 1 hr at 37°C with 100 ng/ml of a range of cytokines. RT-PCR was performed on total RNA as described (Metcalf *et al*, 1995). PCR products were resolved on an agarose gel and Southern blots were hybridised with probes specific for each SOCS family member. Expression of β -actin was assessed to ensure uniformity of amplification.

EXAMPLE 8

10 DNA CONSTRUCTS AND TRANSFECTION

A cDNA encoding epitope-tagged SOCS1 was generated by subcloning the entire SOCS1 coding region into the pEF-BOS expression vector [Mizushima, 1990], engineered to encode an inframe FLAG epitope downstream of an initiation methionine (pF-SOCS1). Using electroporation as described previously [Hilton, 1994], M1 cells expressing the thrombopoietin receptor (M1.mpl) were transfected with the 20 μ g of Aat II-digested pF-SOCS1 expression plasmid and 2 μ g of a Sca I-digested plasmid in which transcription of a cDNA encoding puromycin N-acetyl transferase was driven from the mouse phosphoglycerokinase promoter (pPGKPUropA). After 48 hours in culture, transfected cells were selected with 20 μ g/ml puromycin (Sigma Chemical Company, St Louis MO), and screened for expression of SOCS1 by Western blotting, using the M2 anti-FLAG monoclonal antibody according to the manufacturer's instructions (Eastman Kodak, Rochester NY). In other experiments M1 cells were transfected with only the pF-SOCS1 plasmid or a control and selected by their ability to grow in agar in the presence of 100 ng/ml of IL-6.

25

- 67 -

EXAMPLE 9**IMMUNOPRECIPITATION AND WESTERN BLOTTING**

Prior to either immunoprecipitation or Western blotting, 10^7 M1 cells or their derivatives were washed twice, resuspended in 1ml of DME, and incubated at 37°C for 30 min. The cells were then stimulated for 4 min at 37°C with either saline or 100 ng/ml IL-6, after which sodium vanadate (Sigma Chemical Co., St Louis, MI) was added to a concentration of 1 mM. Cells were placed on ice, washed once with saline containing 1 mM sodium vanadate, and then solubilised for 5 min on ice with 300 µl 1% (v/v) Triton X-100, 150 mM NaCl, 2 mM EDTA, 50 mM Tris-HCl pH 7.4, containing Complete protease inhibitors (Boehringer Mannheim, Mannheim, Germany) and 1 mM sodium vanadate. Lysates were cleared by centrifugation and quantitated using a Coomassie Protein Assay Reagent (Pierce, Rockford IL).

For immunoprecipitations, equal concentrations of protein extracts (1-2 mg) were incubated for 1 hr or overnight at 4°C with either 4 µg of anti-gp130 antibody (M20; Santa Cruz Biotechnology Inc., Santa Cruz, CA) or 4 µg of anti-phosphotyrosine antibody (4G10; Upstate Biotechnology Inc., Lake Placid NY), and 15 µl packed volume of Protein G Sepharose (Pharmacia, Uppsala, Sweden) [Hilton *et al*, 1996]. Immunoprecipitates were washed twice in 1% (v/v) NP40, 150 mM NaCl, 50 mM Tris-HCl pH 8.0, containing Complete protease inhibitors (Boehringer Mannheim, Mannheim, Germany and 1 mM sodium vanadate. The samples were heated for 5 min at 95°C in SDS sample buffer (625 mM Tris-HCl pH 6.8, 0.05% (w/v) SDS, 0.1% (v/v) glycerol, bromophenol blue, 0.125% (v/v) 2-mercaptoethanol), fractionated by SDS-PAGE and immunoblotted as described above.

For Western blotting, 10 µg of protein from a cellular extract or material from an immunoprecipitation reaction was loaded onto 4-15% Ready gels (Bio-Rad Laboratories, Hercules CA), and resolved by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Proteins were transferred to PVDF membrane (Micron Separations Inc., Westborough MA) for 1 hr at 100 V. The membranes were probed with the following primary antibodies; anti-tyrosine phosphorylated STAT3 (1:1000 dilution; New England Biolabs, Beverly, MA); anti-STAT3 (C-20; 1:100 dilution; Santa Cruz Biotechnology Inc., Santa Cruz CA); anti-gp130 (M20, 1:100 dilution; Santa Cruz Biotechnology Inc., Santa Cruz CA); anti-

- 68 -

phosphotyrosine (horseradish peroxidase-conjugated RC20, 1:5000 dilution; Transduction Laboratories, Lexington KY); anti-tyrosine phosphorylated MAP kinase and anti-MAP kinase antibodies (1:1000 dilution; New England Biolabs, Beverly, MA). Blots were visualised using peroxidase-conjugated secondary antibodies and Enhanced Chemiluminescence (ECL) reagents according to the manufacturer's instructions (Pierce, Rockford IL).

EXAMPLE 10

ELECTROPHORETIC MOBILITY SHIFT ASSAYS

Assays were performed as described [Novak, 1995], using the high affinity SIF (c-sis- inducible factor) binding site m67 [Wakao, 1994]. Protein extracts were prepared from M1 cells incubated for 4-10 min at 37°C in 10 ml serum-free DME containing either saline, 100 ng/ml IL-6 or 100 ng/ml IFN- γ . The binding reactions contained 4-6 μ g protein (constant within a given experiment), 5 ng 32 P-labelled m67 oligonucleotide, and 800 ng sonicated salmon sperm DNA. For certain experiments, protein samples were preincubated with an excess of unlabelled m67 oligonucleotide, or antibodies specific for either STAT1 (Transduction Laboratories, Lexington, KY) or STAT3 (Santa Cruz Biotechnology Inc., Santa Cruz CA), as described [Novak, 1995].

Western blots were performed using anti-tyrosine phosphorylated STAT3 or anti-STAT3 (New England Biolabs, Beverly, MA) or anti-gp130 (Santa Cruz Biotechnology Inc.) as described (Nicola *et al*, 1996). EMSA were performed using the m67 oligonucleotide probe, as described (Novak *et al*, 1995).

- 69 -

EXAMPLE 11

EXPRESSION CLONING OF A NOVEL SUPPRESSOR OF CYTOKINE SIGNAL TRANSDUCTION

In order to identify cDNAs capable of suppressing cytokine signal transduction, an expression
5 cloning approach was adopted. This strategy centred on M1 cells, a monocytic leukaemia cell
line that differentiates into mature macrophages and ceases proliferation in response to the
cytokines IL-6, LIF, OSM and IFN- γ , and the steroid dexamethasone. Parental M1 cells were
infected with the RUFneo retrovirus, into which cDNAs from the factor-dependent
haemopoietic cell line FDC-P1 had been cloned. In this retrovirus, transcription of both the
10 neomycin resistance gene and the cloned cDNA was driven off the powerful constitutive
promoter present in the retroviral LTR (Figure 1). When cultured in semi-solid agar, parental
M1 cells form large tightly packed colonies. Upon stimulation with IL-6, M1 cells undergo
rapid differentiation, resulting in the formation in agar of only single macrophages or small
dispersed clusters of cells. Retrovirally-infected M1 cells that were unresponsive to IL-6 were
15 selected in semi-solid agar culture by their ability to form large, tightly packed colonies in the
presence of IL-6 and geneticin. A single stable IL-6-unresponsive clone, 4A2, was obtained
after examining 10^4 infected cells.

A fragment of the neomycin phosphotransferase (neo) gene was used to probe a Southern blot
20 of genomic DNA from clone 4A2 and this revealed that the cell line was infected with a single
retrovirus containing a cDNA approximately 1.4 kbp in length (Figure 2). PCR amplification
using primers from the retroviral vector which flanked the cDNA cloning site enabled recovery
of a 1.4 kbp cDNA insert, which we have named suppressor of cytokine signalling-1, or
SOCS1. This PCR product was used to probe a similar Southern blot of 4A2 genomic DNA
25 and hybridised to two fragments, one which corresponded to the endogenous SOCS1 gene and
the other, which matched the size of the band seen using the neo probe, corresponded to the
SOCS1 cDNA cloned into the integrated retrovirus (Figure 2). The latter was not observed in
an M1 cell clone infected with a retrovirus containing an irrelevant cDNA. Similarly, Northern
blot analysis revealed that SOCS1 mRNA was abundant in the cell line 4A2, but not in the
30 control infected M1 cell clone (Figure 2).

- 70 -

EXAMPLE 12
SOCS1, SOCS2, SOCS3 AND CIS DEFINE A NEW FAMILY
OF SH2-CONTAINING PROTEINS

5 The SOCS1 PCR product was used as a probe to isolate homologous cDNAs from a mouse thymus cDNA library. The sequence of the cDNAs proved to be identical to the PCR product, suggesting that constitutive or over expression, rather than mutation, of the SOCS1 protein was sufficient for generating an IL-6-unresponsive phenotype. Comparison of the sequence of SOCS1 cDNA with nucleotide sequence databases revealed that it was present on mouse and
10 rat genomic DNA clones containing the protamine gene cluster found on mouse chromosome 16. Closer inspection revealed that the 1.4 kb SOCS1 sequence was not homologous to any of the protamine genes, but rather represented a previously unidentified open reading frame located at the extreme 3' end of these clones (Figure 3). There were no regions of discontinuity between the sequences of the SOCS1 cDNA and genomic locus, suggesting that SOCS1 is
15 encoded by a single exon. In addition to the genomic clone containing the protamine genes, a series of murine and human expressed sequenced tags (ESTs) also revealed large blocks of nucleotide sequence identity to mouse SOCS1. The sequence information provided by the human ESTs allowed the rapid cloning of cDNAs encoding human SOCS1.

20 The mouse and rat SOCS1 gene encodes a 212 amino acid protein whereas the human SOCS1 gene encodes a 211 amino acid protein. Mouse, rat and human SOCS1 proteins share 95-99% amino acid identity (Figure 9). A search of translated nucleic acid databases with the predicted amino acid sequence of SOCS1 showed that it was most related to a recently cloned cytokine-inducible immediate early gene product, CIS, and two classes of ESTs. Full length cDNAs
25 from the two classes of ESTs were isolated and found to encode proteins of similar length and overall structure to SOCS1 and CIS. These clones were given the names SOCS2 and SOCS3. Each of the four proteins contains a central SH2 domain and a C-terminal region termed the SOCS motif. The SOCS1 proteins exhibit an extremely high level of amino acid sequence similarity (95-99% identity) amongst different species. However, the forms of the SOCS1,
30 SOCS2, SOCS3 and CIS from the same animal, while clearly defining a new family of SH2-containing proteins, exhibited a lower amino acid identity. SOCS2 and CIS exhibit

- 71 -

approximately 38% amino acid identity, while the remaining members of the family share approximately 25% amino acid identity (Figure 9). The coding region of the genes for SOCS1 and SOC3 appear to contain no introns while the coding region of the genes for SOCS2 and CIS contain one and two introns, respectively.

5

The Genbank Accession Numbers for the sequences referred to herein are mouse SOCS1 cDNA (U88325), human SOCS1 cDNA (U88326), mouse SOCS2 cDNA (U88327), mouse SOCS3 cDNA (U88328).

10

EXAMPLE 13

CONSTITUTIVE EXPRESSION OF SOCS1 SUPPRESSES THE ACTION OF A RANGE OF CYTOKINES

To formally establish that the phenotype of the 4A2 cell line was directly related to expression of SOCS1, and not to unrelated genetic changes which may have occurred independently in
15 these cells, a cDNA encoding an epitope-tagged version of SOCS1 under the control of the EF1 α promoter was transfected into parental M1 cells, and M1 cells expressing the receptor for thrombopoietin, c-mpl (M1.mpl). Transfection of the SOCS1 expression vector into both cell lines resulted in an increase in the frequency of IL-6 unresponsive M1 cells.

20 Multiple independent clones of M1 cells expression SOCS1, as detected by Western blot, displayed a cytokine-unresponsive phenotype that was indistinguishable from 4A2. Further, if transfectants were not maintained in puromycin, expression of SOCS1 was lost over time and cells regained their cytokine responsiveness. In the absence of cytokine, colonies derived from 4A2 and other SOCS1 expressing clones characteristically grew to a smaller size than colonies
25 formed by control M1 cells (Figure 10).

The effect of constitutive SOCS1 expression on the response of M1 cells to a range of cytokines was investigated using the 4A2 cell line and a clone of M1.mpl cells expressing SOCS1 (M1.mpl.SOCS1). Unlike parental M1 cells and M1.mpl cells, the two cell lines
30 expressing SOCS1 continued to proliferate and failed to form differentiated colonies in response to either IL-6, LIF, OSM, IFN- γ or, in the case of the M1.mpl.SOCS1 cell line, thrombopoietin

- 72 -

(Figure 4). For both cell lines, however, a normal response to dexamethasone was observed, suggesting that SOCS1 specifically affected cytokine signal transduction rather than differentiation *per se*. Consistent with these data, while parental M1 cells and M1.mpl cells became large and vacuolated in response to IL-6, 4A2 and M1.mpl.SOCS1 cells showed no
5 evidence of morphological differentiation in response to IL-6 or other cytokines (Figure 5).

EXAMPLE 14

SOCS1 INHIBITS A RANGE OF IL-6 SIGNAL TRANSDUCTION PROCESSES, INCLUDING STAT3 PHOSPHORYLATION

10 AND ACTIVATION

Phosphorylation of the cell surface receptor component gp130, the cytoplasmic tyrosine kinase JAK1 and the transcription factor STAT3 is thought to play a central role in IL-6 signal transduction. These events were compared in the parental M1 and M1.mpl cell lines and their SOCS1-expressing counterparts. As expected, gp130 was phosphorylated rapidly in response
15 to IL-6 in both parental lines, however, this was reduced five- to ten-fold in the cell lines expressing SOCS1 (Figure 6). Likewise, STAT3 phosphorylation was also reduced by approximately ten-fold in response to IL-6 in those cell lines expressing SOCS1 (Figure 6). Consistent with a reduction in STAT3 phosphorylation, activation of specific STAT DNA binding complexes, as determined by electrophoretic mobility shift assay, was also reduced.
20 Notably, there was a reduction in the formation of SIF-A (containing STAT3), SIF-B (STAT1/STAT3 heterodimer) and SIF-C (containing STAT1), the three STAT complexes induced in M1 cells stimulated with IL-6 (Figure 7). Similarly, constitutive expression of SOCS1 also inhibited IFN- γ -stimulated formation of p91 homodimers (Figure 7). STAT phosphorylation and activation were not the only cytoplasmic processes to be effected by
25 SOCS1 expression, as the phosphorylation of other proteins, including shc and MAP kinase, was reduced to a similar extent (Figure 7).

EXAMPLE 15

TRANSCRIPTION OF THE SOCS1 GENE IS STIMULATED BY IL-6

30 *IN VITRO AND IN VIVO*

Although SOCS1 can inhibit cytokine signal transduction when constitutively expressed in M1

- 73 -

cells, this does not necessarily indicate that SOCS1 normally functions to negatively regulate an IL-6 response. In order to investigate this possibility the inventors determined whether transcription of the SOCS1 gene is regulated in the response of M1 cells to IL-6 and, because of the critical role IL-6 plays in regulating the acute phase response to injury and infection, the response of the liver to intravenous injection of 5 mg IL-6. In the absence of IL-6, SOCS1 mRNA was undetectable in either M1 cells or in the liver. However, for both cell types, a 1.4 kb SOCS1 transcript was induced within 20 to 40 minutes by IL-6 (Figure 8). For M1 cells, where the IL-6 was present throughout the experiment, the level of SOCS1 mRNA remained elevated (Figure 8). In contrast, IL-6 was administered in vivo by a single intravenous injection and was rapidly cleared from the circulation, resulting in a pulse of IL-6 stimulation to the liver. Consistent with this, transient expression of SOCS1 mRNA was detectable in the liver, peaking approximately 40 minutes after injection and declining to basal levels within 4 hours (Figure 8).

EXAMPLE 16

REGULATION OF SOCS GENES

Since CIS was cloned as a cytokine-inducible immediate early gene the inventors examined whether SOCS1, SOCS2 and SOCS3 were similarly regulated. The basal pattern of expression of the four SOCS genes was examined by Northern blot analysis of mRNA from a variety of tissues from male and female C57B1/6 mice (Figure 11A). Constitutive expression of SOCS1 was observed in the thymus and to a lesser extent in the spleen and the lung. SOCS2 expression was restricted primarily to the testis and in some animals the liver and lung; for SOCS3 a low level of expression was observed in the lung, spleen and thymus, while CIS expression was more widespread, including the testis, heart, lung, kidney and, in some animals, the liver.

The inventors sought to determine whether expression of the four SOCS genes was regulated by IL-6. Northern blots of mRNA prepared from the livers of untreated and IL-6-injected mice, or from unstimulated and IL-6-stimulated M1 cells, were hybridised with labelled fragments of SOCS1, SOCS2, SOCS3 and CIS cDNAs (Figure 11B). Expression of all four SOCS genes was increased in the liver following IL-6 injection, however the kinetics of

- 74 -

induction appeared to differ. Expression of SOCS1 and SOCS3 was transient in the liver, with mRNA detectable after 20 minutes of IL-6 injection and declining to basal levels within 4 hours for SOCS and 8 hours for SOCS3. Induction of SOCS2 and CIS mRNA in the liver followed similar initial kinetics to that of SOCS1, but was maintained at an elevated level for at least 24 hours. A similar induction of SOCS gene mRNA was observed in other organs, notably the lung and the spleen. In contrast, in M1 cells, while SOCS1 and CIS mRNA were induced by IL-6, no induction of either SOCS2 or SOCS3 expression was detected. This result highlights cell type-specific differences in the expression of the genes of SOCS family members in response to the same cytokine.

10

In order to examine the spectrum of cytokines that was capable of inducing transcription of the various members of the SOCS gene family, bone marrow cells were stimulated for an hour with a range of cytokines, after which mRNA was extracted and cDNA was synthesised. PCR was then used to assess the expression of SOCS1, SOCS2, SOCS3 and CIS (Figure 11C). In the absence of stimulation, little or no expression of any of the SOCS genes was detectable in bone marrow by PCR. Stimulation of bone marrow cells with a broad array of cytokines appeared capable of up regulating mRNA for one or more members of the SOCS family. IFN γ , for example, induced expression of all four SOCS genes, while erythropoietin, granulocyte colony-stimulating factor, granulocyte-macrophage colony stimulating factor and interleukin-3 induced expression of SOCS2, SOCS3 and CIS. Interestingly, tumor necrosis factor alpha, macrophage colony-stimulating factor and interleukin-1, which act through receptors that do not fall into the type I cytokine receptor class also appeared capable of inducing expression of SOCS3 and CIS, suggesting that SOCS proteins may play a broader role in regulating signal transduction.

25 As constitutive expression of SOCS1 inhibited the response of M1 cells to a range of cytokines, the inventors examined whether phosphorylation of the cell surface receptor component gp130 and the transcription factor STAT3, which are thought to play a central role in IL-6 signal transduction, were affected. These events were compared in the parental M1 and M1.mpl cell lines and their SOCS1-expressing counterparts. As expected, gp130 was phosphorylated rapidly in response to IL-6 in both parental lines, however, this was reduced in the cell lines expressing SOCS1 (Figure 12A). Likewise, STAT3 phosphorylation was also reduced in

- 75 -

response to IL-6 in those cell lines expressing SOCS1 (Figure 12A). Consistent with a reduction in STAT3 phosphorylation, activation of specific STAT/DNA binding complexes, as determined by electrophoretic mobility shift assay, was also reduced. Notably, there was a failure to form SIF-A (containing STAT3) and SIF-B(STAT1/STAT3 heterodimer), the major
5 STAT complexes induced in M1 cells stimulated with IL-6 (Figure 12B). Similarly, constitutive expression of SOCS1 also inhibited IFN γ -stimulating formation of SIF-C (STAT1 homodimer; Figure 12B). These experiments are consistent with the proposal that SOCS1 inhibits signal transduction upstream of receptor and STAT phosphorylation, potentially at the level of the JAK kinases.

10

The ability of SOCS1 to inhibit signal transduction and ultimately the biological response to cytokines suggest that, like the SH2-containing phosphatase SHP-1 [Ihle *et al*, 1994; Yi *et al*, 1993], the SOCS proteins may play a central role in controlling the intensity and/or duration of a cell's response to a diverse range of extracellular stimuli by suppressing the signal
15 transduction process. The evidence provided here indicates that the SOCS family acts in a classical negative feedback loop for cytokine signal transduction. Like other genes such as OSM, expression of genes encoding the SOCS proteins is induced by cytokines through the activation of STATs. Once expressed, it is proposed that the SOCS proteins inhibit the activity of JAKs and so reduce the phosphorylation of receptors and STATs, thereby suppressing signal
20 transduction and any ensuing biological response. Importantly, inhibition of STAT activation will, over time, lead to a reduction in SOCS gene expression, allowing cells to regain responsiveness to cytokines.

EXAMPLE 17

25

DATABASE SEARCHES

The NCBI genetic sequence database (Genbank), which encompasses the major database of expressed sequence tags (ESTs) and TIGR database of human expressed sequence tags, were searched for sequences with similarity to a consensus SOCS box sequence using the TFASTA
30 and MOTIF/PATTERN algorithms [Pearson, 1990; Cockwell and Giles, 1989]. Using the software package SRS [Etzold *et al*, 1996], ESTs that exhibited similarity to the SOCS box

- 76 -

(and their partners derived from sequencing the other end of cDNAs) were retrieved and assembled into contigs using Autoassembler (Applied Biosystems, Foster City, CA). Consensus nucleotide sequences derived from overlapping ESTs were then used to search the various databases using BLASTN [Altschul *et al*, 1990]. Again, positive ESTs were retrieved and
 5 added to the contig. This process was repeated until no additional ESTs could be recovered. Final consensus nucleotide sequences were then translated using Sequence Navigator (Applied Biosystems, Foster City, CA).

The ESTs encoding the new SOCS proteins are as follows: **human SOCS4** (EST81149,
 10 EST180909, EST182619, ya99H09, ye70co4, yh53c09, yh77g11, yh87h05, yi45h07, yj04e06, yq12h06, yq56a06, yq60e02, yq92g03, yq97h06, yr90f01, yt69c03, yv30a08, yv55f07, yv57h09, yv87h02, yv98e11, yw68d10, yw82a03, yx08a07, yx72h06, yx76b09, yy37h08, yy66b02, za81f08, zb18f07, zc06e08, zd14g06, zd51h12, zd52b09, ze25g11, ze69f02, zf54f03, zh96e07, zv66h12, zs83a08 and zs83g08). **mouse SOCS-4** (mc65f04, mf42e06, mp10c10,
 15 mr81g09, and mt19h12). **human SOCS-5** (EST15B103, EST15B105, EST27530 and zf50f01). **mouse SOCS-5** (mc55a01, mh98f09, my26h12 and ve24e06). **human SOCS-6** (yf61e08, yf93a09, yg05f12, yg41f04, yg45c02, yh11f10, yh13b05, zc35a12, ze02h08, zl09a03, zl69e10, zn39d08 and zo39e06). **mouse SOCS-6** (mc04c05, md48a03, mf31d03, mh26b07, mh78e11, mh88h09, mh94h07, mi27h04 and mj29c05, mp66g04, mw75g03, va53b05,
 20 vb34h02, vc55d07, vc59e05, vc67d03, vc68d10, vc97h01, vc99c08, vd07h03, vd08c01, vd09b12, vd19b02, vd29a04 and vd46d06). **human SOCS-7** (STS WI30171, EST00939, EST12913, yc29b05, yp49f10, zt10f03 and zx73g04). **mouse SOCS-7** (mj39a01 and vi52h07). **mouse SOCS-8** (mj6e09 and vj27a029). **human SOCS-9** (CSRL-82f2-u, EST114054, yy06b07, yy06g06, zr40c09, zr72h01, yx92c08, yx93b08 and hfe0662). **mouse**
 25 **SOCS-9** (me65d05). **human SOCS-10** (aa48h10, zp35h01, zp97h12, zq08h01, zr34g05, EST73000 and HSDHEI005). **mouse SOCS-10** (mb14d12, mb40f06, mg89b11, mq89e12, mp03g12 and vh53c11). **human SOCS-11** (zt24h06 and zr43b02). **human SOCS-13** (EST59161). **mouse SOCS-13** (ma39a09, me60c05, mi78g05, mk10c11, mo48g12, mp94a01, vb57c07 and vh07c11). **human SOCS-14** (mi75e03, vd29h11 and vd53g07).

30

- 77 -

EXAMPLE 18

cDNA CLONING

Based on the consensus sequences derived from overlapping ESTs, oligonucleotides were
5 designed that were specific to various members of the SOCS family. As described above,
oligonucleotides were labelled and used to screen commercially available genomic and cDNA
libraries cloned with λ bacteriophage. Genomic and/or cDNA clones covering the entire coding
region of mouse SOCS4, mouse SOCS5 and mouse SOCS6 were isolated. The entire gene for
SOCS15 is on the human 12p13 BAC (Genbank Accession Number HSU47924) and the mouse
10 chromosome 6 BAC (Genbank Accession Number AC002393). Partial cDNAs for mouse
SOCS7, SOCS9, SOCS10, SOCS11, SOCS12, SOCS13 and SOCS14 were also isolated.

EXAMPLE 19

NORTHERN BLOTS AND rtPCR

15 Northern blots were performed as described above. The sources of hybridisation probes were
as follows; (i) the entire coding region of the mouse SOCS1 cDNA, (ii) a 1059 bp PCR product
derived from coding region of SOCS5 upstream of the SH2 domain, (iii) the entire coding
region of the mouse SOCS6 cDNA, (iv) a 790 bp PCR product derived from the coding region
20 of a partial SOCS7 cDNA and (v) a 1200 bp Pst I fragment of the chicken glyceraldehyde 3-
phosphate dehydrogenase (GAPDH) cDNA.

EXAMPLE 20

ADDITIONAL MEMBERS OF SOCS FAMILY

25 SOCS1, SOCS2 and SOCS3 are members of the SOCS protein family identified in Examples
1-16. Each contains a central SH2 domain and a conserved motif at the C-terminus, named the
SOCS box. In order to isolate further members of this protein family, various DNA databases
were searched with the amino acid sequence corresponding to conserved residues of the SOCS
30 box. This search revealed the presence of human and mouse ESTs encoding twelve further
members of the SOCS protein family (Figure 13). Using this sequence information cDNAs

- 78 -

encoding SOCS4, SOCS5, SOCS6, SOCS7, SOCS9, SOCS10, SOCS11, SOCS12, SOCS13, SOCS14 and SOCS15 have been isolated. Further analysis of contigs derived from ESTs and cDNAs revealed that the SOCS proteins could be placed into three groups according to their predicted structure N-terminal of the SOCS box. The three groups are those with (i) SH2
5 domains, (ii) WD-40 repeats and (iii) ankyrin repeats.

10

EXAMPLE 21

SOCS PROTEIN WITH SH2 DOMAINS

Eight SOCS proteins with SH2 domains have been identified. These include SOCS1, SOCS2
15 and SOCS3, SOCS5, SOCS9, SOCS11 and SOCS14 (Figure 13). Full length cDNAs were isolated for mouse SOCS5 and SOCS14 and partial clones encoding mouse SOCS9 and SOCS14. Analysis of primary amino acid sequence and genomic structure suggest that pairs of these proteins (SOCS1 and SOCS3, SOCS2 and CIS, SOCS5 and SOCS14 and SOCS9 and SOCS11) are most closely related (Figure 13). Indeed, the SH2 domains of SOCS5 and
20 SOCS14 are almost identical (Figure 13B), and unlike CIS, SOCS1, SOCS2 and SOCS3, SOCS5 and SOCS14 have an extensive, though less well conserved, N-terminal region preceding their SH2 domains (Figure 13A).

EXAMPLE 22

25

SOCS PROTEINS WITH WD-40 REPEATS

Four SOCS proteins with WD-40 repeats were identified. As with the SOCS proteins with SH2 domains, pairs of these proteins appeared to be closely related. Full length cDNAs of mouse SOCS4 and SOCS6 were isolated and shown to encode proteins containing eight WD-
30 40 repeats N-terminal of the SOCS box (Figure 13) and SOCS4 and SOCS6 share 65% amino acid similarity. SOCS15 was recognised as an open reading frame upon sequencing BACs from

- 79 -

human chromosome 12p13 and the syntenic region of mouse chromosome 6 [Ansari-Lari *et al*, 1997]. In the human, chimp and mouse, SOCS15 is encoded by a gene with two coding exons that lies within a few hundred base pairs of the 3' end of the triose phosphate isomerase (TPI) gene, but which is encoded on the opposite strand to TPI (9). In addition to a C-terminal
5 SOCS box, the SOCS15 protein contains four WD-40 repeats. Interestingly, within the EST databases, there is a sequence of a nematode, an insect and a fish relative of SOCS15. SOCS15 appears most closely related to SOCS13.

EXAMPLE 23

10

SOCS PROTEINS WITH ANKYRIN REPEATS

Three SOCS proteins with ankyrin repeats were identified. Analysis of partial cDNAs of mouse SOCS7, SOCS10 and SOCS12 demonstrated the presence of multiple ankyrin repeats.

15

EXAMPLE 24

EXPRESSION PATTERN OF SOCS PROTEINS

The expression of mRNA from representative members of each class of SOCS proteins - SOCS1 and SOCS5 from the SH2 domain group, SOCS6 from the WD-40 repeat group and
20 SOCS7 from the ankyrin repeat group was examined. As shown above, SOCS1 mRNA is found in abundance in the thymus and at lower levels in other adult tissues.

Since transcription of the SOCS1 gene is induced by cytokines, the inventors sought to determine whether levels of SOCS5, SOCS6 and SOCS7 mRNA increased upon cytokine
25 stimulation. In the livers of mice injected with IL-6, SOCS1 mRNA is detectable after 20 min and decreases to background levels within 2 hours. In contrast, the kinetics of SOCS5 mRNA expression are quite different, being only detectable 12 to 24 hours after IL-6 injection. SOCS6 mRNA appears to be expressed constitutively while SOCS7 mRNA was not detected in the liver either before injection of IL-6 or at any time after injection.

30

Expression of these genes was also examined after cytokine stimulation of the factor-dependent

- 80 -

cell line FDCP-1 engineered to express bcl-w. Again, while SOCS6 mRNA was expressed constitutively.

EXAMPLE 25

5

SOCS4

Mouse and human SOCS4 were recognized through searching EST databases using the SOCS box consensus (Figure 13). Those ESTs derived from mouse and human SOCS4 cDNAs are tabulated below (Tables 4.1 and 4.2). Using sequence information derived from mouse ESTs
10 several oligonucleotides were designed and used to screen, in the conventional manner, a mouse thymus cDNA library cloned into λ -bacteriophage. Two cDNAs encoding mouse SOCS4 were isolated and sequenced in their entirety (Figure 15) and shown to overlap the mouse ESTs identified in the database (Table 4.1 and Figure 17). These cDNAs include a region of 5' untranslated region, the entire mouse SOCS4 coding region and a region of 3' untranslated
15 region (Figure 17). Analysis of the sequence confirms that the SOCS4 cDNA encodes a SOCS Box at its C-terminus and a series of 8 WD-40 repeats before the SOCS Box (Figures 17 and 16). The relationship of the two sequence contigs of human SOCS4 (h4.1 and h4.2) to the experimentally determined mouse SOCS4 cDNA sequence is shown in Figure 17. The nucleotide sequence of the two human contigs is listed in Figure 18.

20

SEQ ID NO:13 and 14 represent the nucleotide sequence of murine SOCS4 and the corresponding amino acid sequence. SEQ ID NOs: 15 and 16 are SOCS4 cDNA human contigs h4.1 and h4.2, respectively.

25

EXAMPLE 26

SOCS5

Mouse and human SOCS5 were recognized through searching EST databases using the SOCS box consensus (Figure 13). Those ESTs derived from mouse and human SOCS5 cDNAs are
30 tabulated below (Tables 5.1 and 5.2). Using sequence information derived from mouse and human ESTs, several oligonucleotides were designed and used to screen, in the conventional

- 81 -

manner, a mouse thymus cDNA library, a mouse genomic DNA library and a human thymus cDNA library cloned into λ -bacteriophage. A single genomic DNA clone (57-2) and (5-3-2) cDNA clone encoding mouse SOCS5 were isolated and sequenced in their entirety and shown to overlap with the mouse ESTs identified in the database (Figures 19 and 20A). The entire
5 coding region, in addition to a region of 5' and 3' untranslated regions of mouse SOCS5 appears to be encoded on a single exon (Figure 19). Analysis of the sequence (Figure 20) confirms that SOCS5 genomic and cDNA clones encode a protein with a SOCS box at its C-terminus in addition to an SH2 domain (Figure 19 and 20B). The relationship of the human SOCS5 contig (h5.1; Figure 21) derived from analysis of cDNA clone 5-94-2 and the human
10 SOCS5 ESTs (Table 5.2) to the mouse SOCS5 DNA sequence is shown in Figure 19. The nucleotide sequence and corresponding amino acid sequence of murine SOCS5 are shown in SEQ ID NOs: 17 and 18, respectively. The human SOCS5 nucleotide sequence is shown in SEQ ID NO:19.

15

EXAMPLE 27**SOCS6**

Mouse and human SOCS6 were recognized through searching EST databases using the SOCS box consensus (Figure 13). Those ESTs derived from mouse and human SOCS6 cDNAs are
20 tabulated below (Tables 6.1 and 6.2). Using sequence information derived from mouse ESTs, several oligonucleotides were designed and use to screen, in the conventional manner, a mouse thymus cDNA library. Eight cDNA clones (6-1A, 6-2A, 6-5B, 6-4N, 6-18, 6-29, 6-3N, 6-5N) cDNA clone encoding mouse SOCS6 were isolated and sequenced in their entirety and shown to overlap with the mouse ESTs identified in the database (Figures 22 and 23A). Analysis of
25 the sequence (Figure 23) confirms that the mouse SOCS6 cDNA clones encode a protein with a SOCS box at its C-terminus in addition to a eight WD-40 repeats (Figures 22 and 23B). The relationship of the human SOCS-6 contigs (h6.1 and h6.2 ; Figure 24) derived from analysis of human SOCS6 ESTs (Table 6.2) to the mouse SOCS6 DNA sequence is shown in Figure 22. The nucleotide and corresponding amino acid sequences of murine SOCS6 are shown in SEQ
30 ID NOs: 20 and 21, respectively. SOCS6 human contigs h6.1 and h6.2 are shown in SEQ ID NOs: 22 and 23, respectively.

- 82 -

EXAMPLE 28**SOCS7**

Mouse and human SOCS7 were recognized through searching EST databases using the SOCS
5 box consensus (Figure 13). Those ESTs derived from mouse and human SOCS-7 cDNAs are
tabulated below (Tables 7.1 and 7.2). Using sequence information derived from mouse ESTs,
several oligonucleotides were designed and use to screen, in the conventional manner, a mouse
thymus cDNA library. One cDNA clone (74-10A-11) cDNA clone encoding mouse SOCS7
was isolated and sequenced in its entirety and shown to overlap with the mouse ESTs identified
10 in the database (Figures 25 and 26A). Analysis of the sequence (Figure 26) suggests that
mouse SOCS7 encodes a protein with a SOCS box at its C-terminus, in addition to several
ankyrin repeats (Figure 25 and 26B). The relationship of the human SOCS7 contigs (h7.1 and
h7.2 ; Figure 27) derived from analysis of human SOCS7 ESTs (Table 7.2) to the mouse
SOCS7 DNA sequence is shown in Figure 25. The nucleotide and corresponding amino acid
15 sequences of murine SOCS7 are shown in SEQ ID NOs: 24 and 25, respectively. The
nucleotide sequence of SOCS7 human contigs h7.1 and h7.2 are shown in SEQ ID NOs: 26 and
27, respectively.

EXAMPLE 29

20

SOCS8

ESTs derived from mouse SOCS8 cDNAs are tabulated below (Table 8.1). As described for
other members of the SOCS family, it is possible to isolate cDNAs for mouse SOCS8 using
sequence information derived from mouse ESTs. The relationship of the ESTs to the predicted
25 coding region of SOCS8 is shown in Figure 28. With the nucleotide sequence obtained from
the ESTs shown in Figure 29A and the partial amino acid sequence of SOCS8 shown in Figure
29B. The nucleotide sequence and corresponding amino acid sequences for murine SOCS8 are
shown in SEQ ID NOs:28 and 29, respectively.

30

EXAMPLE 30**SOCS9**

- 83 -

Mouse and human SOCS-9 were recognized through searching EST databases using the SOCS box consensus (Figure 13). Those ESTs derived from mouse and human SOCS9 cDNAs are tabulated below (Tables 9.1 and 9.2). The relationship of the mouse SOCS9 contigs (m9.1; Figure 9.2) derived from analysis of the mouse SOCS9 EST (Table 9.1) to the human SOCS-9 DNA contig (h9.1; Figure 32) derived from analysis of human SOCS9 ESTs (Table 9.2) is shown in Figure 31. Analysis of the sequence (Figure 32) indicates that the human SOCS9 cDNA encodes a protein with a SOCS box at its C-terminus, in addition to an SH2 domain (Figure 30). The nucleotide sequence of murine SOCS9 cDNA is shown in SEQ ID NO:30. The nucleotide sequence of human SOCS9 cDNA is shown in SEQ ID NO:31.

10

EXAMPLE 31

SOCS10

Mouse and human SOCS10 were recognized through searching EST databases using the SOCS box consensus (Figure 13). Those ESTs derived from mouse and human SOCS10 cDNAs are tabulated below (Table 10.1 and 10.2). Using sequence information derived from mouse ESTs, several oligonucleotides were designed and used to screen, in the conventional manner, a mouse thymus cDNA library. Four cDNA clones (10-9, 10-12, 10-23 and 10-24) encoding mouse SOCS10 were isolated, sequenced in their entirety and shown to overlap with the mouse and human ESTs identified in the database (Figures 33 and 34). Analysis of the sequence (Figure 34) indicates that the mouse SOCS10 cDNA clone is not full length but that it does encode a protein with a SOCS box at its C-terminus, in addition to several ankyrin repeats (Figure 33). The relationship of the human SOCS10 contigs (h10.1 and h10.2 ; Figure 35) derived from analysis of human SOCS10 ESTs (Table 10.2) to the mouse SOCS10 DNA sequence is shown in Figure 33. Comparison of mouse cDNA clones and ESTs with human ESTs suggests that the 3' untranslated regions of mouse and human SOCS10 differ significantly. The nucleotide sequence of murine SOCS10 is shown in SEQ ID NO:32 and the nucleotide sequence of SOCS10 human contigs h10.1 and h10.2 are shown in SEQ ID NOs:33 and 34, respectively.

30

EXAMPLE 32

SOCS11

- 84 -

Human SOCS11 were recognized through searching EST databases using the SOCS box consensus (Figure 13). Those ESTs derived from human SOCS11 cDNAs are tabulated below (Table 11.1 and 11.2). The relationship of the human SOCS11 contigs (h11.1; Figure 36A, B), derived from analysis ESTs (Table 11.2) to the predicted encoded protein, is shown in Figure 37. Analysis of the sequence indicates that the human SOCS11 cDNA encodes a protein with a SOCS box at its C-terminus, in addition to an SH2 domain (Figure 37 and 36B). The nucleotide sequence and corresponding amino acid sequence of human SOCS11 are represented in SEQ ID NOs:35 and 36, respectively.

EXAMPLE 33

SOCS12

Mouse and human SOCS-12 were recognized through searching EST databases using the SOCS box consensus (Figure 13). Those ESTs derived from mouse and human SOCS12 cDNAs are tabulated below (Tables 12.1 and 12.2). Using sequence information derived from mouse ESTs, several oligonucleotides were designed and used to screen, in the conventional manner, a mouse thymus cDNA library. Four cDNA clones (10-9, 10-12, 10-23 and 10-24) encoding mouse SOCS12 were isolated, sequenced in their entirety and shown to overlap with the mouse and human ESTs identified in the database (Figures 38 and 39). Analysis of the sequence (Figure 39 and 40) indicates that the SOCS12 cDNA clone encodes a protein with a SOCS box at its C-terminus, in addition to several ankyrin repeats (Figure 38). The relationship of the human SOCS12 contigs (h12.1 and h12.2 ; Figure 40) derived from analysis of human SOCS12 ESTs (Table 12.2) to the mouse SOCS12 DNA sequence is shown in Figure 38. Comparison of mouse cDNA clones and ESTs with human ESTs suggests that the 3' untranslated regions of mouse and human SOCS12 differ significantly. The nucleotide sequence of SOCS12 is shown in SEQ ID NO:37. The nucleotide sequence of human SOCS12 contigs h12.1 and h12.2 are shown in SEQ ID NOs:38 and 39, respectively.

EXAMPLE 34

SOCS13

- 85 -

Mouse and human SOCS-13 were recognized through searching EST databases using the SOCS box consensus (Figure 13). Those ESTs derived from mouse and human SOCS13 cDNAs are tabulated below (Tables 13.1 and 13.2). Using sequence information derived from mouse ESTs, several oligonucleotides were designed and use to screen, in the conventional
5 manner, a mouse thymus and a mouse embryo cDNA library. Three cDNA clones (62-1, 62-6-7 and 62-14) encoding mouse SOCS13 were isolated, sequenced in their entirety and shown to overlap with the mouse ESTs identified in the database (Figure 41 and 42A). Analysis of the sequence (Figure 42) indicates that the mouse SOCS13 cDNA encodes a protein with a SOCS box at its C-terminus, in addition to a potential WD-40 repeat (Figure 41 and 42B). The
10 relationship of the human SOCS13 contigs (h13.1 and h13.2 ; Figure 43) derived from analysis of human SOCS13 ESTs (Table 13.2) to the mouse SOCS13 DNA sequence is shown in Figure 41. The nucleotide sequence and corresponding amino acid sequence of murine SOCS13 and shown in SEQ ID NOs:40 and 41, respectively. The nucleotide sequence of human SOCS13 contig h13.1 is shown in SEQ ID NO:42.

15

EXAMPLE 35

SOCS14

Mouse and human SOCS-14 were recognized through searching EST databases using the
20 SOCS box consensus (Figure 13). Those ESTs derived from mouse and human SOCS14 cDNAs are tabulated below (Tables 14.1 and 14.2). Using sequence information derived from mouse and human ESTs, several oligonucleotides were designed and use to screen, in the conventional manner, a mouse thymus cDNA library, a mouse genomic DNA library and a human thymus cDNA library cloned into λ -bacteriophage . A single genomic DNA clone (57-
25 2) and (5-3-2) cDNA clone encoding mouse SOCS14 were isolated and sequenced in their entirety and shown to overlap with the mouse ESTs identified in the database (Figures 44 and 45A). The entire coding region, in addition to a region of 5' and 3' untranslated regions, of mouse SOCS14 appears to be encoded on a single exon (Figure 44). Analysis of the sequence (Figure 45) confirms that SOCS14 genomic and cDNA clones encode a protein with a SOCS
30 box at its C-terminus in addition to an SH2 domain (Figure 44 and 45B). The relationship of the human SOCS14 contig (h14.1; Figure 14.3) derived from analysis of cDNA clone 5-94-2

- 86 -

and the human SOCS14 ESTs (Table 14.2) to the mouse SOCS14 DNA sequence is shown in Figure 44.

The nucleotide sequence and corresponding amino acid sequence of murine SOCS14 are
5 shown in SEQ ID NOs: 43 and 44, respectively.

- 87 -

EXAMPLE 36**SOCS15**

Mouse and human SOCS15 were recognized through searching DNA databases using the
5 SOCS box consensus (Figure 13). Those ESTs derived from mouse and human SOCS15
cDNAs are tabulated below (Tables 15.1 and 15.2), as are a mouse and human BAC that
contain the entire mouse and human SOCS-15 genes. Using sequence information derived from
the ESTs and the BACs it is possible to predict the entire amino acid sequence of SOCS15 and
as described for the other SOCS genes it is feasible to design specific oligonucleotide probes
10 to allow cDNAs to be isolated. The relationship of the BACs to the ESTs is shown in Figure
46 and the nucleotide and predicted amino acid sequence of the SOCS-15, derived from the
mouse and human BACs is shown in Figures 47 and 48. The nucleotide sequence and
corresponding amino acid sequence of murine SOCS15 are shown in SEQ ID NOs:46 and 47,
respectively. The nucleotide and corresponding amino acid sequence of human SOCS15 are
15 shown in SEQ ID NO:48 and 49, respectively.

EXAMPLE 37**SOCS INTERACTION WITH JAK2 KINASE**

20 These Examples show interaction between SOCS and JAK2 kinase. Interaction is mediated *via*
the SH2 domain of SOCS1, 2, 3 and CIS. The interaction resulted in inhibition of JAK2 kinase
activity by SOCS1 (Figure 49). General interaction between JAK2 and SOCS1, 2, 3, and CIS
is shown in Figure 50.

25 The following methods are employed:

Immunoprecipitation: Cos 6 cells were transiently transfected by electroporation and cultured
for 48 hours. Cells were then lysed on ice in lysis buffer (50 mM Tris/HCL, pH 7.5, 150 mM
NaCl, 1% v/v Triton-X-100, 1 mM EDTA, 1 mM Naf, 1 mM Na₃VO₄) with the addition of
30 complete protease inhibitors (Boehringer Mannheim), centrifuged at 4°C (14,000 x g, 10 min)
and the supernatant retained for immunoprecipitation. JAK2 proteins were immunoprecipitated

- 88 -

using 5 μ l anti-JAK2 antibody (UBI). Antigen-antibody complexes were recovered using protein A-Sepharose (30 μ l of a 50% slurry).

Western blotting: Immunoprecipitates were analysed by sodium dodecyl sulphate (SDS) -
5 polyacrylamide gel electrophoresis (PAGE) under reducing conditions. Protein was then
electrophoretically transferred to nitrocellulose, blocked overnight in 10% w/v skim-milk and
washed in PBS/0.1% v/v Tween-20 (Sigma) (wash buffer) prior to incubation with either anti-
phosphotyrosine antibody (4G10) (1:5000, UBI), anti-FLAG antibody (1.6 μ g/ml) or anti-JAK2
antibody (1:2000, UBI) diluted in wash buffer/1% w/v BSA for 2 hr. Nitrocellulose blots were
10 washed and primary antibody detected with either peroxidase-conjugated sheep anti-rabbit
immunoglobulin (1:5000, Silenus) or peroxidase-conjugated sheep anti-mouse immunoglobulin
(1:5000, Silenus) diluted in wash buffer/1% w/v BSA. Blots were washed and antibody binding
visualised using the enhanced chemiluminescence (ECL) system (Amersham, UK) according
to the manufacturers' instructions.

15

***In-vitro* kinase assay:** An *in vitro* kinase assay was performed to assess intrinsic JAK2 kinase
catalytic activity. JAK2 protein were immunoprecipitated as described, washed twice in kinase
assay buffer (50 mM NaCl, 5 mM MgCl₂, 5 mM MnCl₂, 1 mM NaF, 1 mM Na₃VO₄, 10 mM
HEPES, pH 7.4) and suspended in an equal volume of kinase buffer containing 0.25 μ Ci/ml (γ -
20 ³²P)-ATP (30 min, room temperature). Excess (γ - ³²P)-ATP was removed and the
immunoprecipitates analysed by SDS/PAGE under reducing conditions. Gels were subjected
to a mild alkaline hydrolysis by treatment with 1 M KOH (55°C, 2 hours) to remove
phosphoserine and phosphothreonine. Radioactive bands were visualised with IMAGEQUANT
software on a PhosphorImage system (Molecular Dynamics, Sunnyvale, CA, USA).

25

EXAMPLE 38

MAKING SOCS-1 KNOCKOUT CONSTRUCTS

Diagrams of plasmid constructs and knockout constructs are shown in Figures 51-53. The
30 genomic SOCS-1 clone 95-11-10 was digested with the restriction enzymes BamH1 and EcoR1
to obtain a 3.6Kb DNA fragment 3' of the coding region (SOCS-1 exon), which was used as

- 89 -

the 3' arm in the SOCS-1 knockout vectors. The ends of this fragment were then blunted. This fragment was then ligated into the following vectors:

pBgalpAloxNeo

and pBgalpAloxNeoTK

- 5 which had been linearized at the unique Xho1 site and then blunted. This ligation resulted in the formation of the following vectors:

3'SOCS-1 arm in pBgalpAloxNeo

and 3'SOCS-1 arm in pBgalpAloxNeoTK

- 10 The 5' arm of the SOCS-1 knockout vectors was constructed by using PCR to generate a 2.5Kb PCR product from the genomic SOCS-1 clone 95-11-10 just 5' of the SOCS-1 coding region (SOCS-1 exon). The oligo's used to generate this product were:

5' oligo (sense) (2465)

AGCT AGA TCT GGA CCC TAC AAT GGC AGC [SEQ ID NO:49]

15

3' oligo (antisense) (2466)

AGCT AG ATC TGC CAT CCT ACT CGA GGG GCC AGC TGG [SEQ ID NO:50]

- The PCR product was then digested with the restriction enzyme BglII, to generate BglII ends
20 to the PCR product. This 5' SOCS-1 PCR product, with BglII, ends was then ligated as follows:
3'SOCS-1 arm in pBgalpAloxNeo and 3'SOCS-1 arm in pBgalpAloxNeoTK, which had been
linearized with the unique restriction enzyme BamHI. This resulted in the following vectors
being formed:

5'&3'SOCS-1 arms in pBgalpAloxNeo

- 25 and 5'&3'SOCS-1 arms in pBgalpAloxNeoTK

These were the final SOCS-1 knockout constructs. Both these constructs lacked the entire
SOCS-1 coding region (SOCS-1 EXON), being replaced with portions of the Bgal, B globin
polyA, PGK promoter, neomycin and PGK polyA sequences. The 5'&3'SOCS-1 arms in
30 pBgalpAloxNeoTK vector also contained the thymidine kinase gene sequence, between the
neomycin and PGK poly A sequences.

- 90 -

The vectors: 5'&3'SOCS-1 arms in pBgalpAloxNeo
and 5'&3'SOCS-1 arms in pBgalpAloxNeoTK
were linearized with the unique restriction enzyme NotI and then transfected into Embryonic
5 stem cells by electroporation. Clones which were resistant to neomycin were selected and
analysed by southern blot to determine if they contained the correctly integrated SOCS-1
targeting sequence. In order to determine if correct integration had occurred, genomic DNA
from the neomycin resistant clones was digested with the restriction enzyme EcoRI. The
digested DNA was then blotted onto nylon filters and probed with a 1.5Kb EcoRI /Hind III
10 DNA fragment, which was further 5' of the 5'arm sequence used in the knockout constructs.
The band sizes expected for correct integration were:

Wild type SOCS-1 allele 5.4Kb

15 SOCS-1 knockout allele 8.2Kb in 5'&3'SOCS-1 arms in pBgalpAloxNeo
or 11Kb in 5'&3'SOCS-1 arms in pBgalpAloxNeoTK transfomed cells.

Those skilled in the art will appreciate that the invention described herein is susceptible to
variations and modifications other than those specifically described. It is to be understood that
20 the invention includes all such variations and modifications. The invention also includes all of
the steps, features, compositions and compounds referred to or indicated in this specification,
individually or collectively, and any and all combinations of any two or more of said steps or
features.

- 91 -

Table 4.1**Summary of ESTs derived from mouse SOCS-4 cDNAs**

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-4	Mouse	mc65f04	5'	EST0549700	d13.5-14.5 mouse embryo	m4.1
		mf42e06	5'	EST0593477	d13.5-14.5 mouse embryo	m4.1
		mp10c10	5'	EST0747905	d 8.5 mouse embryo	m4.1
		mr81g09	5'	EST0783081	d13 embryo	m4.1
		mt19h12	5'	EST0816531	spleen	m4.1

Table 4.2**Summary of ESTs derived from human SOCS-4 cDNAs**

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-4	Human	27b5	5'	EST0534081	retina	h4.2
		30d2	5'	EST0534315	retina	h4.2
		J0159F	5'	EST0461188	foetal heart	h4.2
		J3802F	5'	EST0461428	foetal heart	h4.2
		EST19523	5'	EST0958884	retina	h4.2
		EST81149	5'	EST1011015	placenta	h4.2
		EST180909	5'	EST0951375	Jurkat T-lymphocyte	h4.2
		EST182619	5'	EST0953220	Jurkat T-lymphocyte	h4.1
		ya99h09	3'	EST0103262	placenta	h4.2
		ye70c04	5'	EST0172673	foeatl liver/spleen	h4.2
		yh53c09	5'	EST0197390	placenta	h4.2
			3'	EST0197391		h4.2

- 92 -

yh77g11	5'	EST0203418	placenta	h4.2
	3'	EST0203419		h4.1
yh87h05	5'	EST0204888	placenta	h4.1
	3'	EST0204773		h4.1
yi45h07	5'	EST0246604	placenta	h4.2
yj04e06	5'	EST0258541	placenta	h4.1
	3'	EST0258285		h4.1
yq12h06	5'	EST0309968	foetal liver spleen	h4.2
yq56a06	3'	EST0346924	foetal liver spleen	h4.2
yq60e02	5'	EST0347259	foetal liver spleen	h4.2
	3'	EST0347209		h4.2
yq92g03	5'	EST0355932	foetal liver spleen	h4.2
	3'	EST0355884		h4.2
yq97h06	5'	EST0357618	foetal liver spleen	h4.2
	3'	EST0357416		h4.2
yr90f01	5'	EST0372402	foetal liver spleen	h4.2
yt69c03	5'	EST0338395	foetal liver spleen	h4.2
	3'	EST0338303		h4.2
yv30a08	3'	EST0458506	foetal liver spleen	h4.2
yv55f07	5'	EST0465391	foetal liver spleen	h4.2
	3'	EST0463331		h4.2
yv57h09	5'	EST0464336	foetal liver spleen	h4.2
	3'	EST0458765		h4.2
yv87h02	5'	EST0388085	melanocyte	h4.2
yv98e11	5'	EST0400679	melanocyte	h4.2
	3'	EST0400680		h4.2
yw68d10	5'	EST0441370	placenta (8-9 wk)	h4.2
yw82a03	5'	EST0463005	placenta (8-9 wk)	h4.2

- 93 -

	3'	EST0433678		h4.1
yx08a07	3'	EST0407016	melanoocyte	h4.1
yx72h06	5'	EST0435158	melanoocyte	h4.2
	3'	EST0422871	melanoocyte	h4.1
yx76b09	5'	EST0434011	melanoocyte	h4.2
yy37h08	5'	EST0451704	melanoocyte	h4.2
yy66b02	5'	EST0505446	multiple sclerosis lesion	h4.2
za81f08	5'	EST0511777	foetal lung	h4.2
zb18f07	3'	EST0485315	foetal lung	h4.1
zc06e08	5'	EST0540473	parathyroid tumor	h4.1
	3'	EST0540354		h4.1
zd14g06	3'	EST0564666	foetal heart	h4.1
zd51h12	3'	EST0578099	foetal heart	h4.1
zd52b09	5'	EST0582012	foetal heart	h4.1
	3'	EST0581958		h4.1
ze25g11	3'	EST0679543	foetal heart	h4.1
ze69f02	5'	EST0635563	retina	h4.2
	3'	EST0635472		h4.1
zf54f03	5'	EST0680111	retina	h4.2
zh96e07	5'	EST0616241	foetal liver spleen	h4.2
	3'	EST0615745		h4.2
zv66h12	5'	EST1043265	8-9w foetus	h4.2
zs83a08	5'	EST0920072	germinal centre B cell	h4.1
	3'	EST0920016		h4.1
zs83g08	5'	EST0920121	germinal centre B cell	h4.1
	3'	EST0920122		h4.1

- 94 -

Table 5.1
Summary of ESTs derived from mouse SOCS-5 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-5	Mouse	mc55a01	5'	EST0541556	d13.5-14.5 mouse embryo	m5.1
		mh98f09	5'	EST0638237	placenta	m5.1
		my26h12	5'	EST0859939	mixed organs	m5.1
		ve24e06	5'	EST0819106	heart	m5.1

Table 5.2
Summary of ESTs derived from human SOCS-5 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-5	Human	EST15B103	?	EST0258029	adipose tissue	h5.1
		EST15B105	?	EST0258028	adipose tissue	h5.1
		EST27530	5'	EST0965892	cerebellum	h5.1
		zf50f01	5'	EST0679820	retina	h5.1

Table 6.1
Summary of ESTs derived from mouse SOCS-6 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-6	Mouse	mco4c05	5'	EST0525832	d19.5 embryo	m6.1
		md48a03	5'	EST0566730	d13.5-14.5 embryo	m6.1
		mf31d03	5'	EST0675970	d13.5-14.5 embryo	m6.1
		mh26b07	5'	EST0628752	d13.5-14.5 placenta	m6.1
		mh78e11	5'	EST0637608	d13.5-14.5 placenta	m6.1
		mh88h09	5'	EST0644383	d13.5-14.5 placenta	m6.1

- 95 -

mh94h07	5'	EST0638078	d13.5-14.5 placenta	m6.1
mi27h04	5'	EST0644252	d13.5-14.5 embryo	m6.1
mj29c05	5'	EST0664093	d13.5-14.5 embryo	m6.1
mp66g04	5'	EST0757905	thymus	m6.1
mw75g03	5'	EST0847938	liver	m6.1
va53b05	5'	EST0901540	d12.5 embryo	m6.1
vb34h02	5'	EST0930132	lymph node	m6.1
vc55d07	3'	EST1057735	2 cell embryo	m6.1
vc59e05	3'	EST1058201	2 cell embryo	m6.1
vc67d03	3'	EST1057849	2 cell embryo	m6.1
vc68d10	3'	EST1058663	2 cell embryo	m6.1
vc97h01	3'	EST1059343	2 cell embryo	m6.1
vc99c08	3'	EST1059410	2 cell embryo	m6.1
vd07h03	3'	EST1058173	2 cell embryo	m6.1
vd08c01	3'	EST1058275	2 cell embryo	m6.1
vd09b12	3'	EST1058632	2 cell embryo	m6.1
vd19b02	3'	EST1059723	2 cell embryo	m6.1
vd29a04	3'	? none found		m6.1
vd46d06	3'	? none found		m6.1

- 96 -

Table 6.2

Summary of ESTs derived from human SOCS-5 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-6	Human					
		yf61e08	5'	EST0184387	d73 infant brain	h6.1
		yf93a09	5'	EST0186084	d73 infant brain	h6.1
		yg05f12	5'	EST0191486	d73 infant brain	h6.1
		yg41f04	5'	EST0195017	d73 infant brain	h6.1
		yg45c02	5'	EST0185308	d73 infant brain	h6.1
		yh11f10	5'	EST0236705	d73 infant brain	h6.1
		yh13b05	5' 3'	EST0237191 EST0236958	d73 infant brain	h6.1 h6.2
		zc35a12	5'	EST0555518	senescent fibroblasts	h6.1
		ze02h08	5' 3'	EST0603826 EST0603718	foetal heart	h6.1 h6.2
		zl09a03	5' 3'	EST0773936 EST0773892	pregnant uterus	h6.1 h6.1
		zl69e10	5'	EST0683363	colon	h6.1
		zn39d08	5'	EST0718885	endothelial cell	h6.1
		zo39e06	5'	EST0785947	endothelial cell	h6.1

Table 7.1

Summary of ESTs derived from mouse SOCS-7 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-7	Mouse	mj39a01	5'	EST0665627	d13.5/14.5 embryo	m7.1
		vi52h07	5'	EST1267404	d7.5 embryo	m7.1

- 97 -

Table 7.2

Summary of ESTs derived from human SOCS-5 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-7	HUMAN	STS WI-30171		(G21563)	Chromosome 2	h7.2
		EST00939	5'	EST0000906	hippocampus	h7.1
		EST12913	3'	EST0944382	uterus	h7.2
		yc29b05	3'	EST0128727	liver	h7.2
		yp49f10	3'	EST0301914	retina	h7.2
		zt10f03	5'	EST0922932	germinal centre B cell	h7.2
			3'	EST0921231		h7.1
		zx73g04	3'	EST1102975	ovarian tumour	h7.1

Table 8.1

Summary of ESTs derived from mouse SOCS-8 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-8	Mouse	mj16e09	r1	EST0666240	d13.5/14.5 embryo	m8.1
		vj27a029	r1	EST1155973	heart	m8.1

Table 9.1

Summary of ESTs derived from mouse SOCS-9 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
	Mouse	me65d05	5'	EST0585211	d 13.5/14.5 embryo	m9.1

Table 9.2

Summary of ESTs derived from human SOCS-5 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-9	Human	CSRL-83f2-u		(B06659)	chromosome 11	h9.1
		EST114054	5'	EST0939759	placenta	h9.1

- 98 -

yy06b07	3'	EST0434504	melanocyte	h9.1
yy06g06	5'	EST0443783	melanocyte	h9.1
zr40c09	5'	EST0832461	melanocyte, heart, uterus	h9.1
zr72h01	5'	EST0892025	melanocyte, heart, uterus	h9.1
	3'	EST0892026		h9.1
yx92c08	5'	EST0441160	melanocyte	h9.1
yx93b08	5'	EST0441260	melanocyte	h9.1
hfe0662	5'	EST0889611	foetal heart	h9.1

Table 10.1
Summary of ESTs derived from mouse SOCS-10 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
	Mouse	mb14d12	5'	EST0549887	d19.5 embryo	m10.1
		mb40f06	5'	EST0515064	d19.5 embryo	m10.1
		mg89b11	5'	EST0630631	d13.5-14.5 embryo	m10.1
		mq89e12	5'	EST0776015	heart	m10.1
		mp03g12	5'	EST0741991	heart	m10.1
		vh53c11	5'	EST1154634	mammary gland	m10.1

Table 10.2
Summary of ESTs derived from human SOCS-5 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-10	Human	aa48h10	3'	EST1135220	germinal centre B cell	h10.2
		zp35h01	3'	EST0819137	muscle	h10.2
		zp97h12	5'	EST0835442	muscle	h10.2

- 99 -

	3'	EST0831211		h10.2
zq08h01	5'	EST0835907	muscle	h10.1
zr34g05	5'	EST0834251	melanocyte, heart, uterus	h10.2
	3'	EST0834440		h10.2
EST73000	5	EST1004491	ovary	h10.2
HSDHEI005	?	EST0013906	heart	h10.2

Table 11.1
Summary of ESTs derived from human SOCS-5 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-11	Human	zt24h06	r1	EST0925023	ovarian tumor	11.1
		zr43b02	r1	EST0873006	melanocyte, heart, uterus	11.1
			s1	EST0872954		11.1

Table 12.1
Summary of ESTs derived from mouse SOCS-12 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-12	Mouse	EST03803	5'	EST1054173	day 7.5 emb ectoplacental cone	m12.1
		mt18f02	5'	EST0817652	3NbMS spleen	m12.1
		mz60g10	5'	EST0890872	lymph node	m12.1
		va05c11	5'	EST0909449	lymph node	m12.1

Table 12.2
Summary of ESTs derived from human SOCS-5 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-12	Human	STS-SHGC-13867			Chromosome 2	h12.2
		EST177695	5'	EST0948071	Jurkat cells	h12.1

- 100 -

EST64550	5'	EST0997367	Jurkat cells	h12.1
EST76868	5'	EST1007291	pineal body	h12.2
PMY2369	5'	EST1115998	KG-1	h12.1
yb38f04	5'	EST0108807	foetal spleen	h12.1
	3'			h12.2
yg74e12	5'	EST0224407	d73 brain	h12.1
yh13g04	5'	EST0237226	d73 brain	h12.1
	3'	EST0236992		h12.2
yh48b06	5'	yh48b06	placenta	h12.2
yh53a05	5'	EST0197282	placenta	h12.2
	3'	EST0197486		h12.2
yn48h09	5'	EST0278258	brain	h12.2
	3'	EST0278259		h12.2
yn90a09	3'	EST0302557	brain	h12.2
yo08f03	5'	EST0301790	brain	h12.2
	3'	EST0302059		h12.2
yo11e01	3'	? none found		h12.2
yo63b12	5'	EST0303606	breast	h12.2
	3'	EST0304085		h12.2
yq56g02	3'	EST0346935	foetal liver spleen	h12.1
zh57c04	3'	EST0594201	foetal liver spleen	h12.2
zh79h01	3'	EST0598945	foetal liver spleen	h12.2
zh99a11	3'	EST0618570	foetal liver spleen	h12.2
zo92h12	5'	EST0803392	ovarian cancer	h12.1
	3'	EST0803393		h12.2
zs48c01	5'	EST0925714	germinal centre B cell	h12.1
	3'	EST0925530		h12.2

- 101 -

zs45h02	3'	EST0932296	germinal centre B cell	h12.2
---------	----	------------	---------------------------	-------

Table 13.1

Summary of ESTs derived from mouse SOCS-13 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-13	Mouse	ma39c09	5'	EST0517875	day 19.5 embryo	m13.1
		me60c05	5'	EST0584950	day 13.5/14.5 embryo	m13.1
		mi78g05	5'	EST0653834	day 19.5 embryo	m13.1
		mk10c11	5'	EST0735158	day 19.5 embryo	m13.1
		mo48g12	5'	EST0745111	day 10.5 embryo	m13.1
		mp94a01	5'	EST0762827	thymus	m13.1
		vb57c07	5'	EST1028976	day 11.5 embryo	m13.1
		vh07c11	5'	EST1117269	mammary gland	m13.1

Table 13.2

Summary of ESTs derived from human SOCS-13 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-13	Human	EST59161	5'	EST0992726	infant brain	h13.1

Table 14.1

Summary of ESTs derived from mouse SOCS-14 cDNAs

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-14	mouse	mi75e03	5'	EST0651892	d19.5 embryo	m14.1
		vd29h11	5'	EST1067080	2 cell embryo	m14.1

- 102 -

vd53g07	5'	EST1119627	2 cell embryo	m14.1
---------	----	------------	---------------	-------

Table 15.1**Summary of ESTs derived from mouse SOCS-15 cDNAs**

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-15	Mouse	mh29b05	5'	EST0628834	placenta	m15.1
		mh98h09	5'	EST0638243	placenta	m15.1
		ml45a02	5'	EST0687171	testis	m15.1
		mu43a10	5'	EST851588	thymus	m15.1
		my38c09	5'	EST878461	pooled organs	m15.1
		vj37h07	5'	EST1174791	diaphragm	m15.1
		AC002393			Chromosome 6 BAC	m15.1

Table 15.2**Summary of ESTs derived from human SOCS-15 cDNAs**

SOCS	Species	EST name	End	EST no	Library source	Contig
SOCS-15	Human	EST98889	5'	EST1026568	thyroid	h15.1
		ne48bo5	3'	EST1138057	colon tumour	h15.1
		yb12h12	5'	EST0098885	placenta	h15.1
			3'	EST0098886		h15.1
		HSU47924			Chromosome 12 BAC	h15.1

BIBLIOGRAPHY:

Alexander WS, Metcalf D and Dunn AR (1995). *Embo Journal* 14, 5569-78.

Altschul, S.F. Gish, W. Miller, W. Myers, E.W. & Lipman, D.J. (1990) *J. Mol. Biol.* 215, 403-10.

Ansari-Lari, M.A., Shen, Y., Munzy, D.M., Lee, W. and Gibbs, R.A. (1997) *Genome. Res.* 7, 268-280.

Bazan JF (1990). [Review]. *Immunology Today* 11, 350-4.

Bork, P. (1993) *Proteins: Struct. Funct. Genet.* 17, 363-374.

Cockwell, L.Y. and Giles, I.G. (1989) *Comp. Appl. Biosci.* 5, 227-232.

Cutler RL, Liu L, Damen JE and Krystal G (1993). *Journal of Biological Chemistry* 268, 21463-5.

Darnell J Jr., Kerr IM and Stark GR (1994). *Science* 264, 1415-21.

David M, Petricoin E3, Benjamin C, Pine R, Weber MJ and Larner AC (1995). *Science* 269, 1721-3.

David M, Wong L, Flavell R, Thompson SA, Wells A, Larner AC and Johnson GR (1996). *Journal of Biological Chemistry* 271, 9185-8.

Dugaiczky A, Haron JA, Stone EM, Dennison OE, Rothblum KN and Schwartz RJ (1983). *Biochemistry* 22, 1605-13.

Durbin JE, Hackenmiller R, Simon MC and Levy DE (1996). *Cell* 84, 443-50.

- 104 -

Etzold, T., Ulyanov, A. and Argos, P. (1996) *Methods Enzymol.* 266, 114-28.

Gearing DP, Nicola NA, Metcalf D, Foote S, Willson TA, Gough NM and Williams L (1989). *BioTechnology* 7, 1157-1161.

Gupta S, Yan H, Wong LH, Ralph S, Krolewski J and Schindler C (1996). *Embo Journal* 15, 1075-84.

Hilton DJ (1994). An introduction to cytokine receptors, p8-16 in *Guidebook to Cytokines and Their Receptors*, Eds: N. A. Nicola. Oxford University Press: Oxford.

Hilton DJ, Hilton AA, Raicevic A, Rakar S, Harrison-Smith M, Gough NM, Begley CG, Metcalf D, Nicola NA and Willson TA (1994). *Embo Journal* 13, 4765-75.

Hilton DJ, Watowich SS, Katz L and Lodish HF (1996). *J. Biol. Chem.* 271, 4699-4708.

Ichikawa Y (1969). *Journal of Cellular Physiology* 74, 223-34.

Ihle JN (1995). *Nature* 377, 591-4.

Ihle JN, Witthuhn BA, Quelle FW, Yamamoto K and Silvennoinen O (1995). *Annual Review of Immunology* 13, 369-98.

Kaplan MH, Schindler U, Smiley ST and Grusby MJ (1996a). *Immunity* 4, 313-9.

Kaplan MH, Sun YL, Hoey T and Grusby MJ (1996b). *Nature* 382, 174-179.

Levy DE and Stark GR (1996). *Molecular & Cellular Biology* 16, 369-75.

Metcalf D, Wilson TA, Hilton DJ, DiRago L and Mifsud S. (1995) *Leukaemia* 9, 1556-1564.

- 105 -

Meraz MA, White JM, Sheehan KC, Bach EA, Rodig SJ, Dighe AS, Kaplan DH, Riley JK, Greenlund AC, Campbell D, Carver-Moore K, DuBois RN, Clark R, Aguet M and Schreiber RD (1996). *Cell* 84, 431-42.

Mizushima S and Nagata S (1990). *Nucleic Acids Research* 18, 5322.

Murakami M, Narazaki M, Hibi M, Yawata H, Yasukawa K, Hamaguchi M, Taga T and Kishimoto T (1991). *Proc. Natl. Acad. Sci. USA* 88, 11349-11353.

Neer, E.J., Schmidt, C.J., Nambudripad, R. and Smith, T.F. (1994) *Nature* 371, 297-300.

Nicola NA((1994). *Guidebook to Cytokines and Their Receptors*. Oxford University Press: Oxford.

Nicola, NV, Viney E, Hilton DJ, Roberts B and Wilson T. (1996) *Growth Factors* 13, 141-149.

Novak U, Harpur AG, Paradiso L, Kanagasundaram V, Jaworowski A, Wilks AF and Hamilton JA (1995). *Blood* 86, 2948-56.

Pearson WR and Lipman DJ. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444-8.

Pearson WR. (1990) *Methods Enzymol.* 183, 63-98.

Rayner JR and Gonda TJ (1994). *Molecular & Cellular Biology* 14, 880-7.

Sambrook J, Fritsch EF and Maniatis T (1989). *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbour Laboratory Press, Cold Spring Harbour USA.

Sato N, Sakamaki K, Terada N, Arai K and Miyajima A (1993). *Embo Journal* 12, 4181-9.

- 106 -

Shimoda K, van Deursen J, Sangster MY, Sarawar SR, Carson RT, Tripp RA, Chu C, Quelle FW, Nosaka T, Vignali DA, Doherty PC, Grosveld G, Paul WE and Ihle JN (1996). *Nature* 380, 630-3.

Shual K, Ziemiecki A, Wilks AF, Harpur AG, Sadowski HB, Gilman MZ and Darnell JE (1993). *Nature* 366, 580-3.

Sprang SR and Bazan JF (1993). *Curr. Opin. Structural Biol.* 3, 815-827.

Takeda K, Tanaka T, Shi W, Matsumoto M, Minami M, Kashiwamura S, Nakanishi K, Yoshida N, Kishimoto T and Akira S (1996). *Nature* 380, 627-30.

Thierfelder WE, Vandeursen JM, Yamamoto K, Tripp RA, Sarawar SR, Carson RT, Sangster MY, Vignali DDA, Doherty PC, Grosveld GC and Ihle JN (1996). *Nature* 382, 171-174.

Wakao H, Gouilleux F and Groner B (1994). *Embo Journal* 13, 2182-91.

Wen Z, Zhong Z and Darnell J Jr. (1995). *Cell* 82, 241-50.

Yi T, Mui AL, Krystal G and Ihle JN (1993). *Molecular & Cellular Biology* 13, 7577-86.

Yoshimura A, Ohkubo T, Kiguchi T, Jenkins NA, Gilbert DJ, Copeland NG, Hara T and Miyajima A (1995). *Embo Journal* 14, 2816-26.

- 107 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: (Other than US) THE WALTER AND ELIZA HALL INSTITUTE OF
MEDICAL RESEARCH
(US Only)

(ii) TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS

(iii) NUMBER OF SEQUENCES: 49

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: DAVIES COLLISON CAVE
(B) STREET: 1 LITTLE COLLINS STREET
(C) CITY: MELBOURNE
(D) STATE: VICTORIA
(E) COUNTRY: AUSTRALIA
(F) ZIP: 3000

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: PCT INTERNATIONAL
(B) FILING DATE: 31-OCT-1997

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO5117
(B) FILING DATE: 14-FEB-1997

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO 3384
(B) FILING DATE: 01-NOV-1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: HUGHES DR, E JOHN L
(C) REFERENCE/DOCKET NUMBER: EJH/EK

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: +61 3 9254 2777
(B) TELEFAX: +61 3 9254 2770

- 108 -

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CACGCCGCCC ACGTGAAGGC

20

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TTCGCCAATG ACAAGACGCT

20

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1236 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..636

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CGAGGCTCAA GCTCCGGGCG GATTCTGCGT GCCGCTCTCG CTCCTTGGGG TCTGTTGGCC	-101
GGCCTGTGCC ACCCGGACGC CCGGCTCACT GCCTCTGTCT CCCCCATCAG CGCAGCCCCG	-41
GACGCTATGG CCCACCCCTC CAGCTGGCCC CTCGAGTAGG	-1
ATG GTA GCA CGC AAC CAG GTG GCA GCC GAC AAT GCG ATC TCC CCG GCA	48
Met Val Ala Arg Asn Gln Val Ala Ala Asp Asn Ala Ile Ser Pro Ala	
1 5 10 15	
GCA GAG CCC CGA CGG CGG TCA GAG CCC TCC TCG TCC TCG TCT TCG TCC	96
Ala Glu Pro Arg Arg Arg Ser Glu Pro Ser Ser Ser Ser Ser Ser	
20 25 30	
TCG CCA GCG GCC CCC GTG CGT CCC CGG CCC TGC CCG GCG GTC CCA GCC	144

- 109 -

Ser	Pro	Ala	Ala	Pro	Val	Arg	Pro	Arg	Pro	Cys	Pro	Ala	Val	Pro	Ala	
		35					40					45				
CCA	GCC	CCT	GGC	GAC	ACT	CAC	TTC	CGC	ACC	TTC	CGC	TCC	CAC	TCC	GAT	192
Pro	Ala	Pro	Gly	Asp	Thr	His	Phe	Arg	Thr	Phe	Arg	Ser	His	Ser	Asp	
		50				55					60					
TAC	CGG	CGC	ATC	ACG	CGG	ACC	AGC	GCG	CTC	CTG	GAC	GCC	TGC	GGC	TTC	240
Tyr	Arg	Arg	Ile	Thr	Arg	Thr	Ser	Ala	Leu	Leu	Asp	Ala	Cys	Gly	Phe	
		65			70					75					80	
TAT	TGG	GGA	CCC	CTG	AGC	GTG	CAC	GGG	GCG	CAC	GAG	CGG	CTG	CGT	GCC	288
Tyr	Trp	Gly	Pro	Leu	Ser	Val	His	Gly	Ala	His	Glu	Arg	Leu	Arg	Ala	
				85					90					95		
GAG	CCC	GTG	GGC	ACC	TTC	TTG	GTG	CGC	GAC	AGT	CGT	CAA	CGG	AAC	TGC	336
Glu	Pro	Val	Gly	Thr	Phe	Leu	Val	Arg	Asp	Ser	Arg	Gln	Arg	Asn	Cys	
			100					105					110			
TTC	TTC	GCG	CTC	AGC	GTG	AAG	ATG	GCT	TCG	GGC	CCC	ACG	AGC	ATC	CGC	384
Phe	Phe	Ala	Leu	Ser	Val	Lys	Met	Ala	Ser	Gly	Pro	Thr	Ser	Ile	Arg	
		115					120					125				
GTG	CAC	TTC	CAG	GCC	GGC	CGC	TTC	CAC	TTG	GAC	GGC	AGC	CGC	GAG	ACC	432
Val	His	Phe	Gln	Ala	Gly	Arg	Phe	His	Leu	Asp	Gly	Ser	Arg	Glu	Thr	
		130				135					140					
TTC	GAC	TGC	CTT	TTC	GAG	CTG	GAG	CAC	TAC	GTG	GCG	GCG	CCG	CGC		480
Phe	Asp	Cys	Leu	Phe	Glu	Leu	Leu	Glu	His	Tyr	Val	Ala	Ala	Pro	Arg	
		145			150					155					160	
CGC	ATG	TTG	GGG	GCC	CCG	CTG	CGC	CAG	CGC	CGC	GTG	CGG	CCG	CTG	CAG	528
Arg	Met	Leu	Gly	Ala	Pro	Leu	Arg	Gln	Arg	Arg	Val	Arg	Pro	Leu	Gln	
				165					170					175		
GAG	CTG	TGT	CGC	CAG	CGC	ATC	GTG	GCC	GCC	GTG	GGT	CGC	GAG	AAC	CTG	576
Glu	Leu	Cys	Arg	Gln	Arg	Ile	Val	Ala	Ala	Val	Gly	Arg	Glu	Asn	Leu	
			180					185					190			
GCG	CGC	ATC	CCT	CTT	AAC	CCG	GTA	CTC	CGT	GAC	TAC	CTG	AGT	TCC	TTC	624
Ala	Arg	Ile	Pro	Leu	Asn	Pro	Val	Leu	Arg	Asp	Tyr	Leu	Ser	Ser	Phe	
		195					200					205				
CCC	TTC	CAG	ATC	TGA	CCGGCTG	CCGCTGTGCC	GCAGCATTA	GTGGGGGCGC								676
Pro	Phe	Gln	Ile	*												
		210														
CTTATTATTT	CTTATTATTA	ATTATTATTA	TTTTTCTGGA	ACCACGTGGG	AGCCCTCCCC											736
GCCTGGGTCG	GAGGGAGTGG	TTGTGGAGGG	TGAGATGCCT	CCCACTTCTG	GCTGGAGACC											796
TCATCCCACC	TCTCAGGGGT	GGGGGTGCTC	CCCTCCTGGT	GCTCCCTCCG	GGTCCCCCTT											856
GGTTGTAGCA	GCTTGTGTCT	GGGGCCAGGA	CCTGAATTCC	ACTCCTACCT	CTCCATGTTT											916
ACATATTCCC	AGTATCTTTG	CACAAACCAG	GGGTCGGGGA	GGGTCTCTGG	CTTCATTTTT											976
CTGCTGTGCA	GAATATCCTA	TTTTATATTT	TTACAGCCAG	TTAGGTAAT	AAACTTTATT											1036
ATGAAAGTTT	TTTTTTAAAA	GAAAAAAAAA	AAAAAAAAA													1075

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 212 amino acids

- 110 -

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Val Ala Arg Asn Gln Val Ala Ala Asp Asn Ala Ile Ser Pro Ala
 1           5           10           15
Ala Glu Pro Arg Arg Arg Ser Glu Pro Ser Ser Ser Ser Ser Ser
          20           25           30
Ser Pro Ala Ala Pro Val Arg Pro Arg Pro Cys Pro Ala Val Pro Ala
          35           40           45
Pro Ala Pro Gly Asp Thr His Phe Arg Thr Phe Arg Ser His Ser Asp
          50           55           60
Tyr Arg Arg Ile Thr Arg Thr Ser Ala Leu Leu Asp Ala Cys Gly Phe
          65           70           75           80
Tyr Trp Gly Pro Leu Ser Val His Gly Ala His Glu Arg Leu Arg Ala
          85           90           95
Glu Pro Val Gly Thr Phe Leu Val Arg Asp Ser Arg Gln Arg Asn Cys
          100          105          110
Phe Phe Ala Leu Ser Val Lys Met Ala Ser Gly Pro Thr Ser Ile Arg
          115          120          125
Val His Phe Gln Ala Gly Arg Phe His Leu Asp Gly Ser Arg Glu Thr
          130          135          140
Phe Asp Cys Leu Phe Glu Leu Leu Glu His Tyr Val Ala Ala Pro Arg
          145          150          155          160
Arg Met Leu Gly Ala Pro Leu Arg Gln Arg Arg Val Arg Pro Leu Gln
          165          170          175
Glu Leu Cys Arg Gln Arg Ile Val Ala Ala Val Gly Arg Glu Asn Leu
          180          185          190
Ala Arg Ile Pro Leu Asn Pro Val Leu Arg Asp Tyr Leu Ser Ser Phe
          195          200          205
Pro Phe Gln Ile
          210

```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1121 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 223..819

- 111 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCGATCTGTG GGTGACAGTG TCTGCGAGAG ACTTTGCCAC ACCATTCTGC CGGAATTTGG	60
AGAAAAAGAA CCAGCCGCTT CCAGTCCCCT CCCCTCCGC CACCATTTCG GACACCCTGC	120
ACACTCTCGT TTTGGGGTAC CCTGTGACTT CCAGGCAGCA CGCGAGGTCC ACTGGCCCCA	180
GCTCGGGCGA CCAGCTGTCT GGGACGTGTT GACTCATCTC CC ATG ACC CTG CGG	234
Met Thr Leu Arg	
1	
TGC CTG GAG CCC TCC GGG AAT GGA GCG GAC AGG ACG CGG AGC CAG TGG	282
Cys Leu Glu Pro Ser Gly Asn Gly Ala Asp Arg Thr Arg Ser Gln Trp	
5 10 15 20	
GGG ACC GCG GGG TTG CCG GAG GAA CAG TCC CCC GAG GCG GCG CGT CTG	330
Gly Thr Ala Gly Leu Pro Glu Glu Gln Ser Pro Glu Ala Ala Arg Leu	
25 30 35	
GCG AAA GCC CTG CGC GAG CTC AGT CAA ACA GGA TGG TAC TGG GGA AGT	378
Ala Lys Ala Leu Arg Glu Leu Ser Gln Thr Gly Trp Tyr Trp Gly Ser	
40 45 50	
ATG ACT GTT AAT GAA GCC AAA GAG AAA TTA AAA GAG GCT CCA GAA GGA	426
Met Thr Val Asn Glu Ala Lys Glu Lys Leu Lys Glu Ala Pro Glu Gly	
55 60 65	
ACT TTC TTG ATT AGA GAT AGT TCG CAT TCA GAC TAC CTA CTA ACT ATA	474
Thr Phe Leu Ile Arg Asp Ser Ser His Ser Asp Tyr Leu Leu Thr Ile	
70 75 80	
TCC GTT AAG ACG TCA GCT GGA CCG ACT AAC CTG CGG ATT GAG TAC CAA	522
Ser Val Lys Thr Ser Ala Gly Pro Thr Asn Leu Arg Ile Glu Tyr Gln	
85 90 95 100	
GAT GGG AAA TTC AGA TTG GAT TCT ATC ATA TGT GTC AAG TCC AAG CTT	570
Asp Gly Lys Phe Arg Leu Asp Ser Ile Ile Cys Val Lys Ser Lys Leu	
105 110 115	
AAA CAG TTT GAC AGT GTG GTT CAT CTG ATT GAC TAC TAT GTC CAG ATG	618
Lys Gln Phe Asp Ser Val Val His Leu Ile Asp Tyr Tyr Val Gln Met	
120 125 130	
TGC AAG GAT AAA CGG ACA GGC CCA GAA GCC CCA CGG AAT GGG ACT GTT	666
Cys Lys Asp Lys Arg Thr Gly Pro Glu Ala Pro Arg Asn Gly Thr Val	
135 140 145	
CAC CTG TAC CTG ACC AAA CCT CTG TAT ACA TCA GCA CCC ACT CTG CAG	714
His Leu Tyr Leu Thr Lys Pro Leu Tyr Thr Ser Ala Pro Thr Leu Gln	
150 155 160	
CAT TTC TGT CGA CTC GCC ATT AAC AAA TGT ACC GGT ACG ATC TGG GGA	762
His Phe Cys Arg Leu Ala Ile Asn Lys Cys Thr Gly Thr Ile Trp Gly	
165 170 175 180	
CTG CCT TTA CCA ACA AGA CTA AAA GAT TAC TTG GAA GAA TAT AAA TTC	810
Leu Pro Leu Pro Thr Arg Leu Lys Asp Tyr Leu Glu Glu Tyr Lys Phe	
185 190 195	
CAG GTA TAAGTATTTT TCTCTCTTTT TCGTTTTTTT TTAATAAAAA AAAACACAT	866
Gln Val	
GCCTCATATA GACTATCTCC GAATGCAGCT ATGTGAAAGA GAACCCAGAG GCCCTCCTCT	926
GGATAACTGC GCAGAATTCT CTCTTAAGGA CAGTTGGGCT CAGTCTAACT TAAAGGTGTG	986

- 112 -

AAGATGTAGC TAGGTATTTT AAAGTTCCCC TTAGGTAGTT TTAGCTGAAT GATGCTTTCT 1046
 TTCCTATGGC TGCTCAAGAT CAAATGGCCC TTTTAAATGA AACAAAACAA AACAAAACAA 1106
 AAAAAAAAAA AAAAA 1121

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 198 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Thr Leu Arg Cys Leu Glu Pro Ser Gly Asn Gly Ala Asp Arg Thr
 1 5 10 15
 Arg Ser Gln Trp Gly Thr Ala Gly Leu Pro Glu Glu Gln Ser Pro Glu
 20 25 30
 Ala Ala Arg Leu Ala Lys Ala Leu Arg Glu Leu Ser Gln Thr Gly Trp
 35 40 45
 Tyr Trp Gly Ser Met Thr Val Asn Glu Ala Lys Glu Lys Leu Lys Glu
 50 55 60
 Ala Pro Glu Gly Thr Phe Leu Ile Arg Asp Ser Ser His Ser Asp Tyr
 65 70 75 80
 Leu Leu Thr Ile Ser Val Lys Thr Ser Ala Gly Pro Thr Asn Leu Arg
 85 90 95
 Ile Glu Tyr Gln Asp Gly Lys Phe Arg Leu Asp Ser Ile Ile Cys Val
 100 105 110
 Lys Ser Lys Leu Lys Gln Phe Asp Ser Val Val His Leu Ile Asp Tyr
 115 120 125
 Tyr Val Gln Met Cys Lys Asp Lys Arg Thr Gly Pro Glu Ala Pro Arg
 130 135 140
 Asn Gly Thr Val His Leu Tyr Leu Thr Lys Pro Leu Tyr Thr Ser Ala
 145 150 155 160
 Pro Thr Leu Gln His Phe Cys Arg Leu Ala Ile Asn Lys Cys Thr Gly
 165 170 175
 Thr Ile Trp Gly Leu Pro Leu Pro Thr Arg Leu Lys Asp Tyr Leu Glu
 180 185 190
 Glu Tyr Lys Phe Gln Val
 195

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2187 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- 113 -

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 18..695

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGCTGGCTCC GTGCGCC	ATG GTC ACC CAC AGC AAG TTT CCC GCC GCC GGG	50
	Met Val Thr His Ser Lys Phe Pro Ala Ala Gly	
	1 5 10	
ATG AGC CGC CCC CTG GAC ACC AGC CTG CGC CTC AAG ACC TTC AGC TCC	98	
Met Ser Arg Pro Leu Asp Thr Ser Leu Arg Leu Lys Thr Phe Ser Ser		
	15 20 25	
AAA AGC GAG TAC CAG CTG GTG GTG AAC GCC GTG CGC AAG CTG CAG GAG	146	
Lys Ser Glu Tyr Gln Leu Val Val Asn Ala Val Arg Lys Leu Gln Glu		
	30 35 40	
AGC GGA TTC TAC TGG AGC GCC GTG ACC GGC GGC GAG GCG AAC CTG CTG	194	
Ser Gly Phe Tyr Trp Ser Ala Val Thr Gly Gly Glu Ala Asn Leu Leu		
	45 50 55	
CTC AGC GCC GAG CCC GCG GGC ACC TTT CTT ATC CGC GAC AGC TCG GAC	242	
Leu Ser Ala Glu Pro Ala Gly Thr Phe Leu Ile Arg Asp Ser Ser Asp		
	60 65 70 75	
CAG CGC CAC TTC TTC ACG TTG AGC GTC AAG ACC CAG TCG GGG ACC AAG	290	
Gln Arg His Phe Phe Thr Leu Ser Val Lys Thr Gln Ser Gly Thr Lys		
	80 85 90	
AAC CTA CGC ATC CAG TGT GAG GGG GGC AGC TTT TCG CTG CAG AGT GAC	338	
Asn Leu Arg Ile Gln Cys Glu Gly Gly Ser Phe Ser Leu Gln Ser Asp		
	95 100 105	
CCC CGA AGC ACG CAG CCA GTT CCC CGC TTC GAC TGT GTA CTC AAG CTG	386	
Pro Arg Ser Thr Gln Pro Val Pro Arg Phe Asp Cys Val Leu Lys Leu		
	110 115 120	
GTG CAC CAC TAC ATG CCG CCT CCA GGG ACC CCC TCC TTT TCT TTG CCA	434	
Val His His Tyr Met Pro Pro Pro Gly Thr Pro Phe Ser Leu Pro		
	125 130 135	
CCC ACG GAA CCC TCG TCC GAA GTT CCG GAG CAG CCA CCT GCC CAG GCA	482	
Pro Thr Glu Pro Ser Ser Glu Val Pro Glu Gln Pro Pro Ala Gln Ala		
	140 145 150 155	
CTC CCC GGG AGT ACC CCC AAG AGA GCT TAC TAC ATC TAT TCT GGG GGC	530	
Leu Pro Gly Ser Thr Pro Lys Arg Ala Tyr Tyr Ile Tyr Ser Gly Gly		
	160 165 170	
GAG AAG ATT CCG CTG GTA CTG AGC CGA CCT CTC TCC TCC AAC GTG GCC	578	
Glu Lys Ile Pro Leu Val Leu Ser Arg Pro Leu Ser Ser Asn Val Ala		
	175 180 185	
ACC CTC CAG CAT CTT TGT CGG AAG ACT GTC AAC GGC CAC CTG GAC TCC	626	
Thr Leu Gln His Leu Cys Arg Lys Thr Val Asn Gly His Leu Asp Ser		
	190 195 200	
TAT GAG AAA GTG ACC CAG CTG CCT GGA CCC ATT CGG GAG TTC CTG GAT	674	
Tyr Glu Lys Val Thr Gln Leu Pro Gly Pro Ile Arg Glu Phe Leu Asp		
	205 210 215	
CAG TAT GAT GCT CCA CTT TAAGGAGCAA AAGGGTCAGA GGGGGGCCTG	722	
Gln Tyr Asp Ala Pro Leu		
	220 225	

- 114 -

```

GGTCGGTCGG TCGCCTCTCC TCCGAGGCAC ATGGCACAAG CACAAAAATC CAGCCCCAAC      782
GGTCGGTAGC TCCCAGTGAG CCAGGGGCAG ATTGGCTTCT TCCTCAGGCC CTCCACTCCC      842
GCAGAGTAGA GCTGGCAGGA CCTGGAATTC GTCTGAGGGG AGGGGGAGCT GCCACCTGCT      902
TTCCCCCCTC CCCCAGCTCC AGCTTCTTTC AAGTGGAGCC AGCCGGCCTG GCCTGGTGGG      962
ACAATACCTT TGACAAGCGG ACTCTCCCCT CCCCTTCTCT CACACCCCTT CTGCTTCCCA    1022
AGGGAGGTGG GGACACCTCC AAGTGTGTAA CTTAGAACTG CAAGGGGAAT CTTCAAACCT      1082
TCCCGCTGGA ACTTGTTTGC GCTTTGATTT GGTTTGATCA AGAGCAGGCA CCTGGGGGAA    1142
GGATGGAAGA GAAAAGGGTG TGTGAAGGGT TTTTATGCTG GCCAAAGAAA TAACCACTCC      1202
CACTGCCCAA CCTAGGTGAG GAGTGGTGGC TCCTGGCTCT GGGGAGAGTG GCAAGGGGTG      1262
ACCTGAAGAG AGCTATACTG GTGCCAGGCT CCTCTCCATG GGGCAGCTAA TGAAACCTCG      1322
CAGATCCCTT GCACCCCAAG ACCCTCCCCG TTGTGAAGAG GCAGTAGCAT TTAGAAGGGA      1382
GACAGATGAG GCTGGTGAGC TGGCCGCCCT TTCCAACACC GAAGGGAGGC AGATCAACAG      1442
ATGAGCCATC TTGGAGCCCA GGTTTCCCCT GGAGCAGATG GAGGGTTCTG CTTTGTCTCT      1502
CCTATGTGGG GCTAGGAGAC TCGCCTTAAA TGCCCTCTGT CCCAGGGATG GGGATTGGCA      1562
CACAAGGAGC CAAACACAGC CAATAGGCAG AGAGTTGAGG GATTCACCCA GGTGGCTACA      1622
GGCCAGGGGA AGTGGCTGCA GGGGAGAGAC CCAGTCACTC CAGGAGACTC CTGAGTTAAC      1682
ACTGGGAAGA CATTGGCCAG TCCTAGTCAT CTCTCGGTCA GTAGGTCCGA GAGCTTCCAG      1742
GCCCTGCACA GCCCTCCTTT CTCACCTGGG GGGAGGCAGG AGGTGATGGA GAAGCCTTCC      1802
CATGCCGCTC ACAGGGGCCT CACGGGAATG CAGCAGCCAT GCAATTACCT GGAAGTGGTC      1862
CTGTGTTGGG GAGAAACAAG TTTTCTGAAG TCAGGTATGG GGCTGGGTGG GGCAGCTGTG      1922
TGTTGGGGTG GCTTTTTTCT CTCTGTTTTG AATAATGTTT ACAATTTGCC TCAATCACTT      1982
TTATAAAAT CCACCTCCAG CCCGCCCTC TCCCCACTCA GGCTTCGAG GCTGTCTGAA      2042
GATGCTTGAA AACTCAACC AAATCCCAGT TCAACTCAGA CTTTGCACAT ATATTTATAT      2102
TTATACTCAG AAAAGAAACA TTTCAGTAAT TTATAATAAA AGAGCACTAT TTTTAAATGA      2162
AAAAAAAAAA AAAAAAAAAA AAAAAA                                     2187

```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 225 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

Met Val Thr His Ser Lys Phe Pro Ala Ala Gly Met Ser Arg Pro Leu
 1             5             10             15
Asp Thr Ser Leu Arg Leu Lys Thr Phe Ser Ser Lys Ser Glu Tyr Gln
      20             25             30

```

- 115 -

Leu Val Val Asn Ala Val Arg Lys Leu Gln Glu Ser Gly Phe Tyr Trp
 35 40 45
 Ser Ala Val Thr Gly Gly Glu Ala Asn Leu Leu Leu Ser Ala Glu Pro
 50 55 60
 Ala Gly Thr Phe Leu Ile Arg Asp Ser Ser Asp Gln Arg His Phe Phe
 65 70 75 80
 Thr Leu Ser Val Lys Thr Gln Ser Gly Thr Lys Asn Leu Arg Ile Gln
 85 90 95
 Cys Glu Gly Gly Ser Phe Ser Leu Gln Ser Asp Pro Arg Ser Thr Gln
 100 105 110
 Pro Val Pro Arg Phe Asp Cys Val Leu Lys Leu Val His His Tyr Met
 115 120 125
 Pro Pro Pro Gly Thr Pro Ser Phe Ser Leu Pro Pro Thr Glu Pro Ser
 130 135 140
 Ser Glu Val Pro Glu Gln Pro Pro Ala Gln Ala Leu Pro Gly Ser Thr
 145 150 155 160
 Pro Lys Arg Ala Tyr Tyr Ile Tyr Ser Gly Gly Glu Lys Ile Pro Leu
 165 170 175
 Val Leu Ser Arg Pro Leu Ser Ser Asn Val Ala Thr Leu Gln His Leu
 180 185 190
 Cys Arg Lys Thr Val Asn Gly His Leu Asp Ser Tyr Glu Lys Val Thr
 195 200 205
 Gln Leu Pro Gly Pro Ile Arg Glu Phe Leu Asp Gln Tyr Asp Ala Pro
 210 215 220
 Leu
 225

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1094 base pairs
 - (B) TYPE: nucleic acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CTCCGGCTGG CCCCTTCTGT AGGATGGTAG CACACAACCA GGTGGCAGCC GACAATGCAG	60
TCTCCACAGC AGCAGAGCCC CGACGGCGGC CAGAACCTTC CTCCTCTTCC TCCTCCTCGC	120
CCGCGGCCCC CGCGCGCCCG CGGCCGTGCC CCGCGGTCCC GGCCCCGGCC CCCGGCGACA	180
CGCACTTCCG CACATTCCGT TCGCACGCCG ATTACCGGCG CATCACGCGC GCCAGCGCGC	240
TCCTGGACGC CTGCGGATTC TACTGGGGGC CCCTGAGCGT GCACGGGGCG CACGAGCGGC	300
TGCGCGCCGA GCCCGTGGGC ACCTTCCTGG TGCGCGACAG CCGCCAGCGG AACTGCTTTT	360
TCGCCCTTAG CGTGAAGATG GCCTCGGGAC CCACGAGCAT CCGCGTGCAC TTTCAGGCCG	420
GCCGCTTTCA CCTGGATGGC AGCCGCGAGA GCTTCGACTG CCTCTTCGAG CTGCTGGAGC	480

- 116 -

```

ACTACGTGGC GCGCCCGCGC CGCATGCTGG GGGCCCCGCT GCGCCAGCGC CGCGTGCGGC      540
CGCTGCAGGA GCTGTGCCGC CAGCGCATCG TGGCCACCGT GGGCCGCGAG AACCTGGCTC      600
GCATCCCCCT CAACCCCGTC CTCCGCGACT ACCTGAGCTC CTTCCCCTTC CAGATTTGAC      660
CGGCAGCGCC CGCCGTGCAC GCAGCATTAA CTGGGATGCC GTGTTATTTT GTTATTACTT      720
GCCTGGAACC ATGTGGGTAC CCTCCCCGGC CTGGGTGGA GGGAGCGGAT GGGTGTAGGG      780
GCGAGGCGCC TCCCGCCCTC GGCTGGAGAC GAGGCCGCGAG ACCCCTTCTC ACCTCTTGAG      840
GGGGTCCTCC CCCTCCTGGT GCTCCCTCTG GGTCCCCCTG GTTGTGTAG CAGCTTAACT      900
GTATCTGGAG CCAGGACCTG AACTCGCACC TCCTACCTCT TCATGTTTAC ATATACCCAG      960
TATCTTTGCA CAAACCAGGG GTTGGGGGAG GGTCTCTGGC TTTATTTTTC TGCTGTGCAG     1020
AATCCTATTT TATATTTTTC AAAGTCAGTT TAGGTAATAA ACTTTATTAT GAAAGTTTTT     1080
TTTTTTAAAA AAAA                                     1094

```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 211 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```

Met Val Ala His Asn Gln Val Ala Ala Asp Asn Ala Val Ser Thr Ala
1           5           10           15

```

```

Ala Glu Pro Arg Arg Arg Pro Glu Pro Ser Ser Ser Ser Ser Ser Ser
20           25           30

```

```

Pro Ala Ala Pro Ala Arg Pro Arg Pro Cys Pro Ala Val Pro Ala Pro
35           40           45

```

```

Ala Pro Gly Asp Thr His Phe Arg Thr Phe Arg Ser His Ala Asp Tyr
50           55           60

```

```

Arg Arg Ile Thr Arg Ala Ser Ala Leu Leu Asp Ala Cys Gly Phe Tyr
65           70           75           80

```

```

Trp Gly Pro Leu Ser Val His Gly Ala His Glu Arg Leu Arg Ala Glu
85           90           95

```

- 117 -

Pro Val Gly Thr Phe Leu Val Arg Asp Ser Arg Gln Arg Asn Cys Phe
 100 105 110

Phe Ala Leu Ser Val Lys Met Ala Ser Gly Pro Thr Ser Ile Arg Val
 115 120 125

His Phe Gln Ala Gly Arg Phe His Leu Asp Gly Ser Arg Glu Ser Phe
 130 135 140

Asp Cys Leu Phe Glu Leu Leu Glu His Tyr Val Ala Ala Pro Arg Arg
 145 150 155 160

Met Leu Gly Ala Pro Leu Arg Gln Arg Arg Val Arg Pro Leu Gln Glu
 165 170 175

Leu Cys Arg Gln Arg Ile Val Ala Thr Val Gly Arg Glu Asn Leu Ala
 180 185 190

Arg Ile Pro Leu Asn Pro Val Leu Arg Asp Tyr Leu Ser Ser Phe Pro
 195 200 205

Phe Gln Ile
 210

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2807 base pairs
- (B) TYPE: nucleic acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGAAACCGAG GCGGGGAGAC CAGGAGGCCT TGGCCTCAGA GCTTCAGAGT CGCGTGGCAG	60
CAAACAGAGA AACCTGTAGA GGGCAGTGTG CGTCACTTAG CTCAGGGAAG CTGCACGCGA	120
AACTCACCCG CCTTCATTCA TAAACATCGT CAGCTAGGCA CCTACTCCTG GGCTTTCAGG	180
ACAAACTGAA TCACGAAACC ACAGTGTCTT TAAAATAGGT CTGACCGCCT GAATCCCTGG	240
CCAAGGTGTG TACGGGGCAT GGGAGCCCTT GTGCAGAGAT GCTTGCAGGA GCCTTGAGGG	300
GCTCTGTAAG ACAGAGGCTA GGAAGACAAA GTTGGGGGCT ACAGCTTCTT GTCCTGCCCG	360
GGGCCTCAGT TTCTTCGGTT GCCCACGTAG GAGTGCAGAG AGTCCAGCCC CTGGGGACCC	420
AACCCAACCC CGCCCAGTTT CCGAGGAACT CGTCCGGGAG CGGGGGCGCC CCTCCCGCAC	480
CGCCTTAGGC TTCCTTTGAA GCCTCTGCGG TCAGGCCACC GCTTCCTGGG AAGCCCAAGC	540

- 118 -

CAAGGCCAGG	CCGAGTGGCC	AACGGGAGGG	GCCCCGCGCG	GATTCTGGAG	GAGGGCGGCG	600
GCCCCACAGG	TCTCCAGGGC	TGGCTAGCCG	GGCTCCTAGA	GCGGAGACTG	CCAAGGCCTT	660
CGGGTCTTGG	GCAGGAAGGA	TCCTGGCAGG	GAGGAGTTGC	TTGGGGGGTG	GGGGGGAAAG	720
GCTCCAGGCG	CGGTGGAGCT	CTGACCAGGA	GAATGCACAC	ACTCGGAGGG	GAGGAGGCGT	780
GTCAGCCCCA	AGCTAGCATC	CCACCCGGGG	AGCAGCGATG	TGGGGCGAAG	GTAGCCAGAG	840
CAAAAGAGCA	GGCACCAGGT	GACACGAAAC	AGAAGATTCC	GGGTAGAGCC	AGAACCCAG	900
AAGTCCCATT	CAGGAAGGT	GCGAGGCGAG	AACGAGTTAG	GTGGACCCTC	TCCAGGGGCA	960
GCCAAAGAAA	TCTAAAGAGA	ACCCGAAGGA	CTTGCCGGAA	AGAGAAACCG	AAAGCGGCGG	1020
TGGGCGGGAT	CGGTGGGCGG	GGCCTCCCTG	GTTTAAGAGC	TTGATGCAGG	GGCGGGCAGC	1080
AGCAGAGAGA	ACTGCGGCCG	TGGCAGCGGC	ACGGCTCCCG	GCCCCGGAGC	ATGCGCGACA	1140
GCAGCCCCGG	AACCCCCAGC	CGCGGCGCCC	CGCGTCCCGC	CGCCAGGTGA	GCCGAGGCAG	1200
CTGCGAAGGA	GCAGGCGGGA	GGGGATGGGA	GGAAGGGGAG	CAGAGCCTGG	CAGGACTATC	1260
CTCGCAGACT	GCATGGCGGG	GTCGTGGATG	CTATGCCTCT	GGCGCCCGCC	CCACCGGCTG	1320
GCCCAGGCGG	CCCCTCGCGC	GCGCGGGGCG	CCGTCAGCCC	CTCCTCTCCG	GCCCTGAGCC	1380
CGGATCGTCC	GCCCGGGTTC	CAGTTCCCGG	CGTGGCCAGT	AGGCGGCAAC	CGCGAGGCGG	1440
CAAGCCACCC	AGCGGGGACG	GCCTGGAGTC	GGGCCCCCTC	CCACGCCCCC	TTCTCCACGC	1500
GCGCGGGGAG	GCAGGGCTCC	ACCGCCAGTC	TGGAAGGGTT	CCACATACAG	GAACGGCCTA	1560
CTTCGCAGAT	GAGCCCACCG	AGGCTCAGGC	TCCGGGCGGA	TTCTGCGTGT	CACCCTCGCT	1620
CCTTGGGGTC	CGCTGGCCGG	CCTGTGCCAC	CCGGACGCCC	GGTTCAGTGC	CTCTGTCTCC	1680
CCCATCAGCG	CAGCCCCGGA	CGCTATGGCC	CACCCCTCCA	GCTGGCCCCCT	CGAGTAGGAT	1740
GGTAGCACGT	AACCAGGTGG	AAGCCGACAA	TGCGATCTCC	CCGGCATCAG	AGCCCCGACG	1800
GCGGCCAGAG	CCATCCTCGT	CCTCGTCTTC	GTCCTCGCCG	GCGGCCCCGG	CGCGTCCCCG	1860
GCCCTGCCCG	GTGGTCCCGG	CCCCGGCTCC	GGGCGACACT	CACCTCCGCA	CC'TCCGCTC	1920
CCACTCTGAT	TACCGGCGCA	TCACGCGGAC	CAGCGCTCTC	CTGGACGCCT	GCGGCTTCTA	1980
CTGGGGACCC	CTGAGCGTGC	ATGGGGCGCA	CGAACGGCTG	CGTTCCGAAC	CCGTGGGCAC	2040
CTTCTTGGTG	CGCGACAGTC	GCCAGCGGAA	CTGCTTCTTC	GCGCTCAGCG	TGAAGATGGC	2100
TTCGGGCCCC	ACGAGCATTG	GTGTGCACTT	CCAGGCCGGC	CGCTTCCACC	TGGACGGCAA	2160
CCGCGAGACC	TTGACTGACC	TCTTCGAGCT	GCTGGAGCAC	TACGTGGCGG	CGCCGCGCCG	2220
CATGTTGGGG	GCCCCACTGC	GCCAGCGCCG	CGTGGCGCCG	CTGCAGGAGC	TGTGTGCGCA	2280
GCGCATCGTG	GCCGCCGTGG	GTCGCGAGAA	CCTGGCACGC	ATCCCTCTTA	ACCCGGTACT	2340
CCGTGACTAC	CTGAGTTCCT	TCCCCTTCCA	GATCTGACCG	GCTGCCGCCG	TGCCCGCAGA	2400
ATTAAGTGGG	AGCGCCTTAT	TATTTCTTAT	TATTAATTAT	TATTATTTTT	CTGGAACCAC	2460
GTGGGAGCCC	TCCCCGCCTA	GGTCGGAGGG	AGTGGGTGTG	GAGGGTGAGA	TCCCTCCCAC	2520
TTCTGGCTGG	AGACCTTATC	CCGCCTCTCG	GGGGGCCTCC	CCTCCTGGTG	CTCCCTCCCG	2580
GTCCCCCTGG	TTGTAGCAGC	TTGTGTCTGG	GGCCAGGACC	TGAACTCCAC	GCCTACCTCT	2640
CCATGTTTAC	ATGTTCCAG	TATCTTTGCA	CAAACCAGGG	GTGGGGGAGG	GTCTCTGGCT	2700
TCATTTTTCT	GCTGTGCAGA	ATATTCTATT	TTATATTTTT	ACATCCAGTT	TAGATAATAA	2760
ACTTTATTAT	GAAAGTTTTT	TTTTTTAAAG	AAACAAAGAT	TTCTAGA		2807

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 212 amino acids

SUBSTITUTE SHEET (RULE 26)

- 119 -

- (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```

Met Val Ala Arg Asn Gln Val Glu Ala Asp Asn Ala Ile Ser Pro Ala
1           5           10           15

Ser Glu Pro Arg Arg Arg Pro Glu Pro Ser Ser Ser Ser Ser Ser Ser
          20           25           30

Ser Pro Ala Ala Pro Ala Arg Pro Arg Pro Cys Pro Val Val Pro Ala
          35           40           45

Pro Ala Pro Gly Asp Thr His Phe Arg Thr Phe Arg Ser His Ser Asp
          50           55           60

Tyr Arg Arg Ile Thr Arg Thr Ser Ala Leu Leu Asp Ala Cys Gly Phe
65           70           75           80

Tyr Trp Gly Pro Leu Ser Val His Gly Ala His Glu Arg Leu Arg Ser
          85           90           95

Glu Pro Val Gly Thr Phe Leu Val Arg Asp Ser Arg Gln Arg Asn Cys
          100          105          110

Phe Phe Ala Leu Ser Val Lys Met Ala Ser Gly Pro Thr Ser Ile Arg
          115          120          125

Val His Phe Gln Ala Gly Arg Phe His Leu Asp Gly Asn Arg Glu Thr
          130          135          140

Phe Asp Cys Leu Phe Glu Leu Leu Glu His Tyr Val Ala Ala Pro Arg
145          150          155          160

Arg Met Leu Gly Ala Pro Leu Arg Gln Arg Arg Val Arg Pro Leu Gln
          165          170          175

Glu Leu Cys Arg Gln Arg Ile Val Ala Ala Val Gly Arg Glu Asn Leu
          180          185          190

```

- 120 -

Ala Arg Ile Pro Leu Asn Pro Val Leu Arg Asp Tyr Leu Ser Ser Phe
 195 200 205

Pro Phe Gln Ile
 210

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1611 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 263..1529

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGAATTCCGG GCGGGCTGTG TGAGTCTGTG AGTGAAGGC GCGCCGGCTC TTTTGTCTGA	60
GTGTGACCCG GTGGCTTTGT TCCAGGCATT CCGGTGATTT CCTCCGGGCA GTCCGCAGAA	120
GCCGCAGCGG CCGCCCGCGC TCTCTCTGCA GTCTCCACAC CCGGGAGAGC CTGAGCCCGC	180
GTCACGCCCC TCAGCCCCCG CTGAGTCCCT TCTCTGTTGT CGCGTCCGAA TCGAGTTCCC	240
GGAATCAGAC GGTGCCCCAT AG ATG GCC AGC TTT CCC CCG AGG GTT AAC GAG	292
Met Ala Ser Phe Pro Pro Arg Val Asn Glu	
1 5 10	
AAA GAG ATC GTG AGA TCA CGT ACT ATA GGG GAA CTC TTG GCT CCA GCA	340
Lys Glu Ile Val Arg Ser Arg Thr Ile Gly Glu Leu Leu Ala Pro Ala	
15 20 25	
GCT CCT TTT GAC AAG AAA TGT GGT GGT GAG AAC TGG ACG GTT GCT TTT	388
Ala Pro Phe Asp Lys Lys Cys Gly Gly Glu Asn Trp Thr Val Ala Phe	
30 35 40	

- 121 -

GCT CCT GAT GGT TCC TAC TTT GCG TGG TCA CAA GGA TAT CGC ATA GTG	436
Ala Pro Asp Gly Ser Tyr Phe Ala Trp Ser Gln Gly Tyr Arg Ile Val	
45 50 55	
AAG CTT GTC CCG TGG TCC CAG TGC CGT AAG AAC TTT CTT TTG CAT GGT	484
Lys Leu Val Pro Trp Ser Gln Cys Arg Lys Asn Phe Leu Leu His Gly	
60 65 70	
TCC AAA AAT GTT ACC AAT TCA AGC TGT CTA AAA TTG GCA AGA CAA AAC	532
Ser Lys Asn Val Thr Asn Ser Ser Cys Leu Lys Leu Ala Arg Gln Asn	
75 80 85 90	
AGT AAT GGT GGT CAG AAA AAC AAG CCT CCT GAG CAC GTT ATA GAC TGT	580
Ser Asn Gly Gly Gln Lys Asn Lys Pro Pro Glu His Val Ile Asp Cys	
95 100 105	
GGA GAC ATA GTC TGG AGT CTT GCT TTT GGG TCT TCA GTT CCA GAA AAA	628
Gly Asp Ile Val Trp Ser Leu Ala Phe Gly Ser Ser Val Pro Glu Lys	
110 115 120	
CAG AGT CGT TGC GTT AAT ATA GAA TGG CAT CGG TTC CGA TTT GGA CAG	676
Gln Ser Arg Cys Val Asn Ile Glu Trp His Arg Phe Arg Phe Gly Gln	
125 130 135	
GAT CAG CTA CTC CTT GCC ACA GGA TTA AAC AAT GGT CGC ATC AAA ATC	724
Asp Gln Leu Leu Leu Ala Thr Gly Leu Asn Asn Gly Arg Ile Lys Ile	
140 145 150	
TGG GAT GTA TAT ACA GGA AAA CTC CTC CTT AAT TTG GTA GAC CAC ATT	772
Trp Asp Val Tyr Thr Gly Lys Leu Leu Leu Asn Leu Val Asp His Ile	
155 160 165 170	
GAA ATG GTT AGA GAT TTA ACT TTT GCT CCA GAT GGG AGC TTA CTC CTT	820
Glu Met Val Arg Asp Leu Thr Phe Ala Pro Asp Gly Ser Leu Leu Leu	
175 180 185	
GTA TCA GCT TCA AGA GAC AAA ACT CTA AGA GTG TGG GAC CTG AAA GAT	868
Val Ser Ala Ser Arg Asp Lys Thr Leu Arg Val Trp Asp Leu Lys Asp	
190 195 200	
GAT GGA AAC ATG GTG AAA GTA TTG CGG GCA CAT CAG AAT TGG GTG TAC	916
Asp Gly Asn Met Val Lys Val Leu Arg Ala His Gln Asn Trp Val Tyr	
205 210 215	
AGT TGT GCA TTC TCT CCC GAC TGT TCT ATG CTG TGT TCA GTG GGC GCC	964

- 122 -

Ser Cys Ala Phe Ser Pro Asp Cys Ser Met Leu Cys Ser Val Gly Ala	
220 225 230	
AGT AAA GCA GTT TTC CTT TGG AAT ATG GAT AAA TAC ACC ATG ATT AGG	1012
Ser Lys Ala Val Phe Leu Trp Asn Met Asp Lys Tyr Thr Met Ile Arg	
235 240 245 250	
AAG CTG GAA GGT CAT CAC CAT GAT GTT GTA GCT TGT GAC TTT TCT CCT	1060
Lys Leu Glu Gly His His His Asp Val Val Ala Cys Asp Phe Ser Pro	
255 260 265	
GAT GGA GCA TTG CTA GCT ACT GCA TCC TAT GAC ACT CGT GTG TAT GTC	1108
Asp Gly Ala Leu Leu Ala Thr Ala Ser Tyr Asp Thr Arg Val Tyr Val	
270 275 280	
TGG GAT CCA CAC AAT GGA GAC CTT CTG ATG GAG TTT GGG CAC CTG TTT	1156
Trp Asp Pro His Asn Gly Asp Leu Leu Met Glu Phe Gly His Leu Phe	
285 290 295	
CCC TCG CCC ACT CCA ATA TTT GCT GGA GGA GCA AAT GAC CGA TGG GTG	1204
Pro Ser Pro Thr Pro Ile Phe Ala Gly Gly Ala Asn Asp Arg Trp Val	
300 305 310	
AGA GCT GTG TCT TTC AGT CAT GAT GGA CTG CAT GTT GCC AGC CTT GCT	1252
Arg Ala Val Ser Phe Ser His Asp Gly Leu His Val Ala Ser Leu Ala	
315 320 325 330	
GAT GAT AAA ATG GTG AGG TTC TGG AGA ATC GAT GAG GAT TGT CCG GTA	1300
Asp Asp Lys Met Val Arg Phe Trp Arg Ile Asp Glu Asp Cys Pro Val	
335 340 345	
CAA GTT GCA CCT TTG AGC AAT GGT CTT TGC TGT GCC TTT TCT ACT GAT	1348
Gln Val Ala Pro Leu Ser Asn Gly Leu Cys Cys Ala Phe Ser Thr Asp	
350 355 360	
GGC AGT GTT TTA GCT GCT GGG ACA CAT GAT GGA AGT GTG TAT TTT TGG	1396
Gly Ser Val Leu Ala Ala Gly Thr His Asp Gly Ser Val Tyr Phe Trp	
365 370 375	
GCC ACT CCA AGG CAA GTC CCT AGC CTT CAA CAT ATA TGT CGC ATG TCA	1444
Ala Thr Pro Arg Gln Val Pro Ser Leu Gln His Ile Cys Arg Met Ser	
380 385 390	
ATC CGA AGA GTG ATG TCC ACC CAA GAA GTC CAA AAA CTG CCT GTT CCT	1492
Ile Arg Arg Val Met Ser Thr Gln Glu Val Gln Lys Leu Pro Val Pro	

- 123 -

395	400	405	410	
TCC AAA ATA TTG GCG TTT CTC TCC TAC CGC GGT TAG A CTGAAGACTG				1539
Ser Lys Ile Leu Ala Phe Leu Ser Tyr Arg Gly *				
	415	420		
CCTTTCCTGG TAGGCCTGCC AGACAGAGCG CCCTTTACAA GACACACCTC AAGCTTTACC				1599
TCGTGCCGAA TT				1611

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 422 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met	Ala	Ser	Phe	Pro	Pro	Arg	Val	Asn	Glu	Lys	Glu	Ile	Val	Arg	Ser
1				5				10						15	
Arg	Thr	Ile	Gly	Glu	Leu	Leu	Ala	Pro	Ala	Ala	Pro	Phe	Asp	Lys	Lys
		20					25						30		
Cys	Gly	Gly	Glu	Asn	Trp	Thr	Val	Ala	Phe	Ala	Pro	Asp	Gly	Ser	Tyr
		35					40						45		
Phe	Ala	Trp	Ser	Gln	Gly	Tyr	Arg	Ile	Val	Lys	Leu	Val	Pro	Trp	Ser
		50					55						60		
Gln	Cys	Arg	Lys	Asn	Phe	Leu	Leu	His	Gly	Ser	Lys	Asn	Val	Thr	Asn
		65				70				75					80
Ser	Ser	Cys	Leu	Lys	Leu	Ala	Arg	Gln	Asn	Ser	Asn	Gly	Gly	Gln	Lys
			85						90					95	
Asn	Lys	Pro	Pro	Glu	His	Val	Ile	Asp	Cys	Gly	Asp	Ile	Val	Trp	Ser
			100					105						110	
Leu	Ala	Phe	Gly	Ser	Ser	Val	Pro	Glu	Lys	Gln	Ser	Arg	Cys	Val	Asn
			115					120						125	

- 124 -

Ile	Glu	Trp	His	Arg	Phe	Arg	Phe	Gly	Gln	Asp	Gln	Leu	Leu	Leu	Ala	130	135	140	
Thr	Gly	Leu	Asn	Asn	Gly	Arg	Ile	Lys	Ile	Trp	Asp	Val	Tyr	Thr	Gly	145	150	155	160
Lys	Leu	Leu	Leu	Asn	Leu	Val	Asp	His	Ile	Glu	Met	Val	Arg	Asp	Leu	165	170	175	
Thr	Phe	Ala	Pro	Asp	Gly	Ser	Leu	Leu	Leu	Val	Ser	Ala	Ser	Arg	Asp	180	185	190	
Lys	Thr	Leu	Arg	Val	Trp	Asp	Leu	Lys	Asp	Asp	Gly	Asn	Met	Val	Lys	195	200	205	
Val	Leu	Arg	Ala	His	Gln	Asn	Trp	Val	Tyr	Ser	Cys	Ala	Phe	Ser	Pro	210	215	220	
Asp	Cys	Ser	Met	Leu	Cys	Ser	Val	Gly	Ala	Ser	Lys	Ala	Val	Phe	Leu	225	230	235	240
Trp	Asn	Met	Asp	Lys	Tyr	Thr	Met	Ile	Arg	Lys	Leu	Glu	Gly	His	His	245	250	255	
His	Asp	Val	Val	Ala	Cys	Asp	Phe	Ser	Pro	Asp	Gly	Ala	Leu	Leu	Ala	260	265	270	
Thr	Ala	Ser	Tyr	Asp	Thr	Arg	Val	Tyr	Val	Trp	Asp	Pro	His	Asn	Gly	275	280	285	
Asp	Leu	Leu	Met	Glu	Phe	Gly	His	Leu	Phe	Pro	Ser	Pro	Thr	Pro	Ile	290	295	300	
Phe	Ala	Gly	Gly	Ala	Asn	Asp	Arg	Trp	Val	Arg	Ala	Val	Ser	Phe	Ser	305	310	315	320
His	Asp	Gly	Leu	His	Val	Ala	Ser	Leu	Ala	Asp	Asp	Lys	Met	Val	Arg	325	330	335	
Phe	Trp	Arg	Ile	Asp	Glu	Asp	Cys	Pro	Val	Gln	Val	Ala	Pro	Leu	Ser	340	345	350	
Asn	Gly	Leu	Cys	Cys	Ala	Phe	Ser	Thr	Asp	Gly	Ser	Val	Leu	Ala	Ala	355	360	365	

- 125 -

Gly Thr His Asp Gly Ser Val Tyr Phe Trp Ala Thr Pro Arg Gln Val
 370 375 380

Pro Ser Leu Gln His Ile Cys Arg Met Ser Ile Arg Arg Val Met Ser
 385 390 395 400

Thr Gln Glu Val Gln Lys Leu Pro Val Pro Ser Lys Ile Leu Ala Phe
 405 410 415

Leu Ser Tyr Arg Gly *
 420

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 783 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTGTCTTCCT CCGCAGCGCG AGGCTGGGTA CAGGGTCTAT TGTCTGTGGT TGA CTCCGTA	60
CTTTGGTCTG AGGCCTTCGG GAGCTTTCCC GAGGCAGTTA GCAGAAGCCG CAGCGACCGC	120
CCCCGCCCCG CTCTCTGTGC CCTGGGCCCCG GGAGACAAAC TTGGCGTCAC GCCCTCAGCG	180
GTCGCCACTC TCTTCTCTGT TGTGGGTCC GCATCGTATT CCCGGAATCA GACGGTGCCC	240
CATAGATGGC CAGCTTTCCC CCGAGGGTCA ACGAGAAAGA GATCGTGAGA TCACGTACTA	300
TAGGTGAACT TTTAGCTCCT GCAGCTCCTT TTGACAAGAA ATGTGGTCGT GAAAATTGGA	360
CTGTTGCTTT TGCTCCAGAT GGTTCATACT TTGCTTGGTC ACAAGGACAT CGCACAGTAA	420
AGCTTGTTCC GTGGTCCCAG TGCCTTCAGA ACTTTCTCTT GCATGGCACC AAGAATGTTA	480
CCAATTCAAG CAGTTTAAGA TTGCCAAGAC AAAATAGTGA TGGTGGTCAG AAAAATAAGC	540
CTCGTGACAT ATTATAGACT GTGGAGATAT AGTCTGGAGT CTTGCTTTTG GGTCA TCAGT	600

- 126 -

TCCAGAAAAA CAGAGTCGCT GTGTAAATAT AGAATGGCAT CGCTTCAGAT TTGGACAAGA	660
TCAGCTACTT CTTGCTACAG GGTGAACAA TGGGCGTATC AAAATATGGG ATGTATATCA	720
GGAAACTCCT CCTTAACTTG GTAGATCATA CTGAAGTGGT CAGAGATTTA ACTTTTGCTC	780
CAG	783

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1122 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTCTGTATGT CTGAATGAAG CTATAACATT TGCCTTTTTTA TTGCAGGTTT TCCTTTGGAA	60
TATGGATAAA TACACCATGA TACGGAAACT AGAAGGACAT CACCATGATG TGGTAGCTTG	120
TGACTTTTCT CCTGATGGAG CATTACTGGC TACTGCATCT TATGATACTC GAGTATATAT	180
CTGGGATCCA CATAATGGAG ACATTCTGAT GGAATTTGGG CACCTGTTTC CCCCACCTAC	240
TCCAATATTT GCTGGAGGAG CAAATGACCG GTGGGTACGA TCTGTATCTT TTAGCCATGA	300
TGGACTGCAT GTTGCAAGCC TTGCTGATGA TAAAATGGTG AGGTTCTGGA GAATTGATGA	360
GGATTATCCA GTGCAAGTTG CACCTTTGAG CAATGGTCTT TGCTGTGCCT TCTCTACTGA	420
TGGCAGTGTT TTAGCTGCTG GGACACATGA CGGAAGTGTG TATTTTGGG CCACTCCACG	480
GCAGGTCCCT AGCCTGCAAC ATTTATGTCG CATGTCAATC CGAAGAGTGA TGCCCACCCA	540
AGAAGTTCAG GAGCTGCCGA TTCCTTCCAA GCTTTTGGAG TTTCTCTCGT ATCGTATTTA	600
GAAGATTCTG CTTCCCTAG TAGTAGGGAC TGACAGAATA CACTTAACAC AAACCTCAAG	660
CTTTACTGAC TTCAATTATC TGTTTTTAAA GACGTAGAAG ATTTATTTAA TTTGATATGT	720

- 127 -

```

TCTTGACTG CATTTTGATC AGTTGAGCTT TTAAATATT ATTTATAGAC AATAGAAGTA      780
TTTCTGAACA TATCAATAT AAATTTTTTT AAAGATCTAA CTGTGAAAAC ATACATACCT      840
GTACATATTT AGATATAAGC TGCTATATGT TGAATGGACC CTTTGTCTTT TCTGATTTTT      900
AGTTCTGACA TGTATATATT GCTTCAGTAG AGCCACAATA TGTATCTTTG CTGTAAAGTG      960
CAAGGAAATT TTAAATCTG GGACACTGAG TTAGATGGTA AATACTGACT TACGAAAGTT     1020
GAATTGGGTG AGGCGGGCAA ATCACCTGAG GTCAGCAGTT TGAGACTAGC CTGGCAAACA     1080
TGATGAAACC CTGTCTCTAC TAAAAATACA AAAAAAAAAA AA                        1122

```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2537 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 422..2029

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```

CGGCACGAGC CGGGCTCCGT CCGGAGGAAG CGAGGCTGCG CCGCCGGCCC GGCAGGAGCG      60
GAGGACGGGA GCGCGGGCGG TCGCGCTCGC CCTGTCGCTG ACTGCGCTGC CCCGGCCCAT      120
CCTTGCCTGG CCGCAGGTGC CCTGGATGAG GCCGCCGCGC GTGTCCCGGC CGCTGAGTGT      180
CCCCCGCGGT CGCCCGGCGC CTGCCCTCAA GCGGCCGCCT CTCCTTGCCC GGGTCCCCGT      240
TTTCCCCCGG CGCAGTCCTC CTCCGGTGGG CGCCTCCGCA CCTCGGCGCA GCGGGCACGG      300
CCCTCGGGCC GGGATGGATC CGCCGGGAAG AGGAAGACAA GCCGGGGCGT TGAGCCCCTG      360

```

- 128 -

CGCACGGTGC CGCCGCGCGT AGTGGGAGCT TACTCGCAGT AGGCTCTCGC TCTTCTAATC	420
A ATG GAT AAA GTG GGG AAA ATG TGG AAC AAC TTA AAA TAC AGA TGC	466
Met Asp Lys Val Gly Lys Met Trp Asn Asn Leu Lys Tyr Arg Cys	
1 5 10 15	
CAG AAT CTC TTC AGC CAC GAG GGA GGA AGC CGT AAT GAG AAC GTG GAG	514
Gln Asn Leu Phe Ser His Glu Gly Gly Ser Arg Asn Glu Asn Val Glu	
20 25 30	
ATG AAC CCC AAC AGA TGT CCG TCT GTC AAA GAG AAA AGC ATC AGT CTG	562
Met Asn Pro Asn Arg Cys Pro Ser Val Lys Glu Lys Ser Ile Ser Leu	
35 40 45	
GGA GAG GCA GCT CCC CAG CAA GAG AGC AGT CCC TTA AGA GAA AAT GTT	610
Gly Glu Ala Ala Pro Gln Gln Glu Ser Ser Pro Leu Arg Glu Asn Val	
50 55 60	
GCC TTA CAG CTG GGA CTG AGC CCT TCC AAG ACC TTT TCC AGG CGG AAC	658
Ala Leu Gln Leu Gly Leu Ser Pro Ser Lys Thr Phe Ser Arg Arg Asn	
65 70 75	
CAA AAC TGT GCC GCA GAG ATC CCT CAA GTG GTT GAA ATC AGC ATC GAG	706
Gln Asn Cys Ala Ala Glu Ile Pro Gln Val Val Glu Ile Ser Ile Glu	
80 85 90 95	
AAA GAC AGT GAC TCG GGT GCC ACC CCA GGA ACG AGG CTT GCA CGG AGA	754
Lys Asp Ser Asp Ser Gly Ala Thr Pro Gly Thr Arg Leu Ala Arg Arg	
100 105 110	
GAC TCC TAC TCG CGG CAC GCC CCG TGG GGA GGA AAG AAG AAA CAT TCC	802
Asp Ser Tyr Ser Arg His Ala Pro Trp Gly Gly Lys Lys Lys His Ser	
115 120 125	
TGT TCC ACA AAG ACC CAG AGT TCA TTG GAT ACC GAG AAA AAG TTT GGT	850
Cys Ser Thr Lys Thr Gln Ser Ser Leu Asp Thr Glu Lys Lys Phe Gly	
130 135 140	
AGA ACT CGA AGC GGC CTT CAG AGG CGA GAG CGG CGC TAT GGA GTC AGC	898
Arg Thr Arg Ser Gly Leu Gln Arg Arg Glu Arg Arg Tyr Gly Val Ser	
145 150 155	
TCC ATG CAG GAC ATG GAC AGC GTT TCT AGC CGC GCG GTC GGG AGC CGC	946
Ser Met Gln Asp Met Asp Ser Val Ser Ser Arg Ala Val Gly Ser Arg	
160 165 170 175	

- 129 -

TCC CTG AGG CAG AGG CTC CAG GAC ACG GTG GGT TTG TGT TTT CCC ATG	994
Ser Leu Arg Gln Arg Leu Gln Asp Thr Val Gly Leu Cys Phe Pro Met	
180 185 190	
AGA ACT TAC AGC AAG CAG TCA AAG CCA CTC TTT TCC AAT AAA AGA AAA	1042
Arg Thr Tyr Ser Lys Gln Ser Lys Pro Leu Phe Ser Asn Lys Arg Lys	
195 200 205	
ATA CAT CTT TCT GAA TTA ATG CTG GAG AAA TGC CCT TTT CCT GCT GGC	1090
Ile His Leu Ser Glu Leu Met Leu Glu Lys Cys Pro Phe Pro Ala Gly	
210 215 220	
TCG GAT TTA GCA CAA AAG TGG CAT TTG ATT AAA CAG CAT ACC GCC CCT	1138
Ser Asp Leu Ala Gln Lys Trp His Leu Ile Lys Gln His Thr Ala Pro	
225 230 235	
GTG AGC CCA CAC TCA ACA TTT TTT GAT ACA TTT GAT CCA TCA CTG GTG	1186
Val Ser Pro His Ser Thr Phe Phe Asp Thr Phe Asp Pro Ser Leu Val	
240 245 250 255	
TCT ACA GAA GAT GAA GAA GAT AGG CTT CGC GAG AGA AGA CGG CTT AGT	1234
Ser Thr Glu Asp Glu Glu Asp Arg Leu Arg Glu Arg Arg Arg Leu Ser	
260 265 270	
ATC GAA GAA GGG GTG GAT CCC CCT CCC AAC GCA CAA ATA CAC ACC TTT	1282
Ile Glu Glu Gly Val Asp Pro Pro Pro Asn Ala Gln Ile His Thr Phe	
275 280 285	
GAA GCT ACT GCA CAG GTC AAC CCA TTG TAT AAG CTG GGA CCA AAG TTA	1330
Glu Ala Thr Ala Gln Val Asn Pro Leu Tyr Lys Leu Gly Pro Lys Leu	
290 295 300	
GCT CCT GGG ATG ACA GAG ATA AGT GGA GAT GGT TCT GCA ATT CCA CAA	1378
Ala Pro Gly Met Thr Glu Ile Ser Gly Asp Gly Ser Ala Ile Pro Gln	
305 310 315	
GCA ATT GTG ACT CAG AAG AGG ATT CAA CCA CCC TAT GTC TGC AGT CAC	1426
Ala Ile Val Thr Gln Lys Arg Ile Gln Pro Pro Tyr Val Cys Ser His	
320 325 330 335	
GGA GGC AGA AGC AGC GCC AGG TGT CCG GGG ACA GCC ACG CGC ACG TTA	1474
Gly Gly Arg Ser Ser Ala Arg Cys Pro Gly Thr Ala Thr Arg Thr Leu	
340 345 350	
GCA GAC AGG GAG CTT GGA AAG TTC ATA CGC AGA TCG ATT ACA TAC ACT	1522

- 130 -

Ala Asp Arg Glu Leu Gly Lys Phe Ile Arg Arg Ser Ile Thr Tyr Thr	
355 360 365	
GCC TCG TGC CAG ATT TGC TTC AGA TCA CAG GGA ATC CCT GTT ACT GGG	1570
Ala Ser Cys Gln Ile Cys Phe Arg Ser Gln Gly Ile Pro Val Thr Gly	
370 375 380	
GCG TGA TGG ACC GAT ACG AGG CCG AAG CCC TTC TAG AAG GGA AAC CGG	1618
Ala * Trp Thr Asp Thr Arg Pro Lys Pro Phe * Lys Gly Asn Arg	
385 390 395	
AAG GCA CGT TCT TGC TCA GGG ACT CTG CAC AGG AGG ACT ACC TCT TCT	1666
Lys Ala Arg Ser Cys Ser Gly Thr Leu His Arg Arg Thr Thr Ser Ser	
400 405 410 415	
CTG TGA GCT TCC GCC GCT ACA ACA GGT CTC TGC ACG CCC GGA TCG AGC	1714
Leu * Ala Ser Ala Ala Thr Thr Gly Leu Cys Thr Pro Gly Ser Ser	
420 425 430	
AGT GGA ACC ACA ACT TCA GCT TCG ATG CCC ATG ACC CCT GCG TGT TTC	1762
Ser Gly Thr Thr Thr Ser Ala Ser Met Pro Met Thr Pro Ala Cys Phe	
435 440 445	
ACT CCT CCA CGT CAC GGG GCT TCT CGA ACA CTA TAA AGA CCC CAG CTC	1810
Thr Pro Pro Arg His Gly Ala Ser Arg Thr Leu * Arg Pro Gln Leu	
450 455 460	
TTG CAT GTT TTT TGA ACC GTT GCT AAC GAT ATC ACT GAA TAG AAC TTT	1858
Leu His Val Phe * Thr Val Ala Asn Asp Ile Thr Glu * Asn Phe	
465 470 475	
CCC TTT CAG CCT GCA GTA TAT CTG CCG CGC AGT GAT CTG CAG ATG CAC	1906
Pro Phe Gln Pro Ala Val Tyr Leu Pro Arg Ser Asp Leu Gln Met His	
480 485 490 495	
TAC GTA TGA TGG GAT TGA CGG GCT CCC GCT ACC GTC GAT GTT ACA GGA	1954
Tyr Val * Trp Asp * Arg Ala Pro Ala Thr Val Asp Val Thr Gly	
500 505 510	
TTT TTT AAA AGA GTA TCA TTA TAA ACA AAA AGT TAG GGT TCG CTG GTT	2002
Phe Phe Lys Arg Val Ser Leu * Thr Lys Ser * Gly Ser Leu Val	
515 520 525	
AGA ACG AGA CCA GTC AAA GCA AAG TAACTCCTGT CCCCAAAGGG CACTAACTAA	2056
Arg Thr Arg Pro Val Lys Ala Lys	

530

535

GTCTGCTCCT	CCCGTGCATC	GAAC TGCACC	CATAGGAGGC	AGTCAGCTGC	TAGGATTTTC	2116
CACCCAGAAT	GGGAGCTTAG	TCATTAGCCT	CTGCCCTATG	GGGTCCGCTG	TTCCTCAGAC	2176
AAAGGTGCCT	AGGGACAGCA	AGATGGCTTG	CAGGTGTTTCG	GTGGGCTGTG	ACAACTGAGG	2236
GAGGCAACTC	TGGGGCATTT	GCTATGAAGA	ATTCTATTTC	TTACCGAAGA	ACAAATTATT	2296
AATATTGGAT	GGGTATTTCA	ATAGTGTGAC	TAATGTTTGA	AATTATTTTT	TCTAAGAATT	2356
TTTCTATAAC	CTTCAGAAAA	AGTAGTGATG	TTTGTAGTTA	CTATAAATCA	AGCTTTGAAA	2416
GTTCAAAACA	AACAAGTTAA	ATAAAAAGACT	ACCTTCCTTT	TAGAGAAAAC	AAATGCAAGT	2476
TTTCCCAGCC	ACAGGCATTG	TGCACTGTTA	ATGTTGCTTG	TTATCAGCTC	CTTTCTCCTC	2536
C						2537

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 535 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Asp Lys Val Gly Lys Met Trp Asn Asn Leu Lys Tyr Arg Cys Gln
1 5 10 15

Asn Leu Phe Ser His Glu Gly Gly Ser Arg Asn Glu Asn Val Glu Met
20 25 30

Asn Pro Asn Arg Cys Pro Ser Val Lys Glu Lys Ser Ile Ser Leu Gly

35 40 45

Glu Ala Ala Pro Gln Gln Glu Ser Ser Pro Leu Arg Glu Asn Val Ala
50 55 60

Leu Gln Leu Gly Leu Ser Pro Ser Lys Thr Phe Ser Arg Arg Asn Gln

- 132 -

65		70		75		80
Asn Cys Ala Ala Glu Ile Pro Gln Val Val Glu Ile Ser Ile Glu Lys						
	85		90		95	
Asp Ser Asp Ser Gly Ala Thr Pro Gly Thr Arg Leu Ala Arg Arg Asp						
	100		105		110	
Ser Tyr Ser Arg His Ala Pro Trp Gly Gly Lys Lys Lys His Ser Cys						
	115		120		125	
Ser Thr Lys Thr Gln Ser Ser Leu Asp Thr Glu Lys Lys Phe Gly Arg						
	130		135		140	
Thr Arg Ser Gly Leu Gln Arg Arg Glu Arg Arg Tyr Gly Val Ser Ser						
	145		150		155	160
Met Gln Asp Met Asp Ser Val Ser Ser Arg Ala Val Gly Ser Arg Ser						
	165		170		175	
Leu Arg Gln Arg Leu Gln Asp Thr Val Gly Leu Cys Phe Pro Met Arg						
	180		185		190	
Thr Tyr Ser Lys Gln Ser Lys Pro Leu Phe Ser Asn Lys Arg Lys Ile						
	195		200		205	
His Leu Ser Glu Leu Met Leu Glu Lys Cys Pro Phe Pro Ala Gly Ser						
	210		215		220	
Asp Leu Ala Gln Lys Trp His Leu Ile Lys Gln His Thr Ala Pro Val						
	225		230		235	240
Ser Pro His Ser Thr Phe Phe Asp Thr Phe Asp Pro Ser Leu Val Ser						
	245		250		255	
Thr Glu Asp Glu Glu Asp Arg Leu Arg Glu Arg Arg Arg Leu Ser Ile						
	260		265		270	
Glu Glu Gly Val Asp Pro Pro Pro Asn Ala Gln Ile His Thr Phe Glu						
	275		280		285	
Ala Thr Ala Gln Val Asn Pro Leu Tyr Lys Leu Gly Pro Lys Leu Ala						
	290		295		300	
Pro Gly Met Thr Glu Ile Ser Gly Asp Gly Ser Ala Ile Pro Gln Ala						

- 133 -

305		310		315		320
Ile Val Thr Gln Lys Arg Ile Gln Pro Pro Tyr Val Cys Ser His Gly						
	325		330		335	
Gly Arg Ser Ser Ala Arg Cys Pro Gly Thr Ala Thr Arg Thr Leu Ala						
	340		345		350	
Asp Arg Glu Leu Gly Lys Phe Ile Arg Arg Ser Ile Thr Tyr Thr Ala						
	355		360		365	
Ser Cys Gln Ile Cys Phe Arg Ser Gln Gly Ile Pro Val Thr Gly Ala						
	370		375		380	
* Trp Thr Asp Thr Arg Pro Lys Pro Phe * Lys Gly Asn Arg Lys						
385		390		395		400
Ala Arg Ser Cys Ser Gly Thr Leu His Arg Arg Thr Thr Ser Ser Leu						
	405		410		415	
* Ala Ser Ala Ala Thr Thr Gly Leu Cys Thr Pro Gly Ser Ser Ser						
	420		425		430	
Gly Thr Thr Thr Ser Ala Ser Met Pro Met Thr Pro Ala Cys Phe Thr						
	435		440		445	
Pro Pro Arg His Gly Ala Ser Arg Thr Leu * Arg Pro Gln Leu Leu						
	450		455		460	
His Val Phe * Thr Val Ala Asn Asp Ile Thr Glu * Asn Phe Pro						
465		470		475		480
Phe Gln Pro Ala Val Tyr Leu Pro Arg Ser Asp Leu Gln Met His Tyr						
	485		490		495	
Val * Trp Asp * Arg Ala Pro Ala Thr Val Asp Val Thr Gly Phe						
	500		505		510	
Phe Lys Arg Val Ser Leu * Thr Lys Ser * Gly Ser Leu Val Arg						
	515		520		525	
Thr Arg Pro Val Lys Ala Lys						
	530		535			

SUBSTITUTE SHEET (RULE 26)

- 134 -

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1221 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GATTAAACAG CATACAGCTC CTGTGAGCCC ACATTCAACA TTTTGTGATA CTTTGATCCA	60
TCTTTGGTTT CTACAGAAGA TGAAGAAGAT AGGCTTAGAG AGAGAAGGCG GCTTAGTATT	120
GAAGAAGGGG TTGATCCCCC TCCAATGCA CAAATACATA CATTTGAAGC TACTGCACAG	180
GTTAATCCAT TATTAACTG GGACCAAAAT TAGCTCCTGG AATGACTGAA ATAAGTGGGG	240
ACAGTTCTGC AATTCCACAA GCTAATTGTG ACTCGGAAGA GGATACAACC ACCCTGTGTT	300
GCAGTCACGG AGGCAGAAGC AGCGTCAGAT ATCTGGAGAC AGCCATACCC ATGTTAGCAG	360
ACAGGGAGCT TGGAAAGTCC ACACACAGAT TGATTACATA CACTGCTTCG TGCCTGATTT	420
GCTTCAAATT ACAGGGAATC CCTGTTACTG GGGAGTGATG GACCGTTATG AAGCAGAAGC	480
CCTTCTCGAA GGGAAACCTG AAGGCACGTT TTTGCTCAGG GACTCTGCGC AAGAGGACTA	540
CTTCTTCTCT GTGAGCTTCC GCCGATACAA CAGATCCCTG CATGCCCCGAA TTGAGCAGTG	600
GAATCACAAC TTTAGTTTCG ACGCCCATGA CCCGTGTGTA TTTCACCTCT CCACTGTAAC	660
GGGACTTTTA GAACATTATA AAGATCCCAG TTCGTGCATG TTTTTTGAAC CATTGCTTAC	720
TATATCACTA AATAGGACTT TCCCTTTTAG CCTGCAGTAT ATCTGTGCGC CGGTAATCTG	780
CAGGTGCACT ACGTATGATG GAATTGATGG GCTCCCTCTA CCCTCAATGT TACAGGATTT	840
TTTAAAAGAG TATCATTATA AACAAAAAGT TAGAGTTCGC TGGTTGGAAC GAGAACCAGT	900
CAAGGCAAAG TAAACTCTCC GGTCCCCAAA GGGTGTTAAC TAGGTCCGCT TTCATGTGCA	960

- 135 -

TCAGACAGTA CACCTATAGC AAGCACACGT AGCAGTGTTA GGCTTTTTTCA TACAGTATGT 1020
 AAGCTTAGTG TTAGTATCTG TCAGATGCTA CCTGCTGTTA CTTATTCAGA TAAACATGGT 1080
 GCCTATTGGA ACAATAGCGG ATAGAGCTAC AGGTGTTTCA TAAGACTACA AAAACATTTT 1140
 GCCTATTTTCG CTAACAGTTT GGTTTTTTAAT GGCTGTGGTA TTTGAGTGAG GCAACTCTGG 1200
 GGCATTTGTT ATGAAGAAAT G 1221

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2369 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 116..1330

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGCACGAGGC GGTGGTGGCG GCGGCGGGCG CGGCCGCGGC GGGCGGGCG CGGAATGAAG 60
 GCCCACGGCC CTGGGGGCTG AGGCGCCCGC CGCCTGGGGC GGGCCGCGCG TCCTC ATG 118
 Met
 1
 GAG GCC GGA GAG GAG CCG CTG CTG CTG GCT GAA CTC AAG CCT GGG CGC 166
 Glu Ala Gly Glu Glu Pro Leu Leu Leu Ala Glu Leu Lys Pro Gly Arg
 5 10 15
 CCC CAC CAG TTC GAC TGG AAG TCA AGC TGC GAG ACC TGG AGC GTG GCC 214
 Pro His Gln Phe Asp Trp Lys Ser Ser Cys Glu Thr Trp Ser Val Ala
 20 25 30
 TTC TCG CCA GAC GGT TCC TGG TTC GCC TGG TCT CAA GGA CAC TGC GTG 262

- 136 -

Phe	Ser	Pro	Asp	Gly	Ser	Trp	Phe	Ala	Trp	Ser	Gln	Gly	His	Cys	Val	
35						40					45					
GTC	AAG	CTG	GTC	CCC	TGG	CCC	TTA	GAG	GAA	CAG	TTC	ATC	CCT	AAA	GGA	310
Val	Lys	Leu	Val	Pro	Trp	Pro	Leu	Glu	Glu	Gln	Phe	Ile	Pro	Lys	Gly	
50					55					60					65	
TTC	GAA	GCC	AAG	AGC	CGA	AGC	AGC	AAG	AAT	GAC	CCA	AAA	GGA	CGG	GGC	358
Phe	Glu	Ala	Lys	Ser	Arg	Ser	Ser	Lys	Asn	Asp	Pro	Lys	Gly	Arg	Gly	
				70					75					80		
AGT	CTG	AAG	GAG	AAG	ACG	CTG	GAC	TGT	GGC	CAG	ATT	GTG	TGG	GGG	CTG	406
Ser	Leu	Lys	Glu	Lys	Thr	Leu	Asp	Cys	Gly	Gln	Ile	Val	Trp	Gly	Leu	
			85					90					95			
GCC	TTC	AGC	CCG	TGG	CCC	TCT	CCA	CCC	AGC	AGG	AAA	CTC	TGG	GCA	CGT	454
Ala	Phe	Ser	Pro	Trp	Pro	Ser	Pro	Pro	Ser	Arg	Lys	Leu	Trp	Ala	Arg	
	100						105					110				
CAC	CAT	CCC	CAG	GCG	CCT	GAT	GTT	TCT	TGC	CTG	ATC	CTG	GCC	ACA	GGT	502
His	His	Pro	Gln	Ala	Pro	Asp	Val	Ser	Cys	Leu	Ile	Leu	Ala	Thr	Gly	
	115					120					125					
CTC	AAC	GAT	GGG	CAG	ATC	AAG	ATT	TGG	GAG	GTA	CAG	ACA	GGC	CTC	CTG	550
Leu	Asn	Asp	Gly	Gln	Ile	Lys	Ile	Trp	Glu	Val	Gln	Thr	Gly	Leu	Leu	
130					135					140					145	
CTT	CTG	AAT	CTT	TCT	GGC	CAC	CAA	GAC	GTC	GTG	AGA	GAT	CTG	AGC	TTC	598
Leu	Leu	Asn	Leu	Ser	Gly	His	Gln	Asp	Val	Val	Arg	Asp	Leu	Ser	Phe	
				150					155						160	
ACG	CCC	AGC	GGC	AGT	TTG	ATT	TTG	GTC	TCT	GCA	TCC	CGG	GAT	AAG	ACA	646
Thr	Pro	Ser	Gly	Ser	Leu	Ile	Leu	Val	Ser	Ala	Ser	Arg	Asp	Lys	Thr	
			165					170					175			
CTT	CGA	ATT	TGG	GAC	CTG	AAT	AAA	CAC	GGT	AAG	CAG	ATC	CAG	GTG	TTA	694
Leu	Arg	Ile	Trp	Asp	Leu	Asn	Lys	His	Gly	Lys	Gln	Ile	Gln	Val	Leu	
	180						185					190				
TCC	GGC	CAT	CTG	CAG	TGG	GTT	TAC	TGC	TGC	TCC	ATC	TCC	CCT	GAC	TGT	742
Ser	Gly	His	Leu	Gln	Trp	Val	Tyr	Cys	Cys	Ser	Ile	Ser	Pro	Asp	Cys	
	195					200					205					
AGC	ATG	CTG	TGC	TCT	GCA	GCT	GGG	GAG	AAG	TCG	GTC	TTT	CTG	TGG	AGC	790
Ser	Met	Leu	Cys	Ser	Ala	Ala	Gly	Glu	Lys	Ser	Val	Phe	Leu	Trp	Ser	

SUBSTITUTE SHEET (RULE 26)

- 137 -

210	215	220	225	
ATG CGG TCC TAC ACA CTA ATC CGG AAA CTA GAA GGC CAC CAA AGC AGT				838
Met Arg Ser Tyr Thr Leu Ile Arg Lys Leu Glu Gly His Gln Ser Ser				
	230	235	240	
GTT GTC TCC TGT GAT TTC TCT CCT GAT TCA GCC TTG CTT GTC ACA GCT				886
Val Val Ser Cys Asp Phe Ser Pro Asp Ser Ala Leu Leu Val Thr Ala				
	245	250	255	
TCG TAT GAC ACC AGT GTG ATT ATG TGG GAC CCC TAC ACC GGC GCG AGG				934
Ser Tyr Asp Thr Ser Val Ile Met Trp Asp Pro Tyr Thr Gly Ala Arg				
	260	265	270	
CTG AGG TCA CTT CAT CAC ACA CAA CTT GAA CCC ACC ATG GAT GAC AGT				982
Leu Arg Ser Leu His His Thr Gln Leu Glu Pro Thr Met Asp Asp Ser				
	275	280	285	
GAC GTC CAC ATG AGC TCC CTG AGG TCC GTG TGC TTC TCA CCT GAA GGC				1030
Asp Val His Met Ser Ser Leu Arg Ser Val Cys Phe Ser Pro Glu Gly				
	290	295	300	305
TTG TAT CTC GCT ACG GTG GCA GAT GAC AGG CTG CTC AGG ATC TGG GCT				1078
Leu Tyr Leu Ala Thr Val Ala Asp Asp Arg Leu Leu Arg Ile Trp Ala				
	310	315	320	
CTG GAA CTG AAG GCT CCG GTT GCC TTT GCT CCG ATG ACC AAT GGT CTT				1126
Leu Glu Leu Lys Ala Pro Val Ala Phe Ala Pro Met Thr Asn Gly Leu				
	325	330	335	
TGC TGC ACG TTC TTC CCA CAC GGT GGA ATT ATT GCC ACA GGG ACG AGA				1174
Cys Cys Thr Phe Phe Pro His Gly Gly Ile Ile Ala Thr Gly Thr Arg				
	340	345	350	
GAT GGC CAT GTC CAG TTC TGG ACA GCT CCC CGG GTC CTG TCC TCA CTG				1222
Asp Gly His Val Gln Phe Trp Thr Ala Pro Arg Val Leu Ser Ser Leu				
	355	360	365	
AAG CAC TTA TGC AGG AAA GCC CTC CGA AGT TTC CTG ACA ACG TAT CAA				1270
Lys His Leu Cys Arg Lys Ala Leu Arg Ser Phe Leu Thr Thr Tyr Gln				
	370	375	380	385
GTC CTA GCA CTG CCA ATC CCC AAG AAG ATG AAA GAG TTC CTC ACA TAC				1318
Val Leu Ala Leu Pro Ile Pro Lys Lys Met Lys Glu Phe Leu Thr Tyr				
	390	395	400	

- 138 -

AGG ACT TTC TAGCAGTGCC GGCTCCCCCA CCTCCTGCAG CAGCAGCAGT	1367
Arg Thr Phe	
405	
ACAAGGGACT GGCTAGGATG GAGTCAGGCA GCTCACACTG GACCAGTGTG GACCTTCCTT	1427
CCTCCCATGG CATGTGCAAG TAGGTCTGCG TGACCCCACT TCTGTGGTGC CGGCCTTACC	1487
TCGTCTTCAT CCGTGGTGAG CAGCCTTCGT CAGTCTAGTT GTGTTGAAGC CAAGTGCAGT	1547
TGTGGATGTT GCTGGGGTAA TAAAGGCAAG CGGGCTCCAG AGCCTCTCTG GTGGCGGCCA	1607
AGCCACACTC CCTTAACTGG GAAGTACCTG CCACGTAGGG CATTTCTGCT GCCTATTTCC	1667
AGCCAGCGGC TGCATGGTTT GAAGTTCCTC CGTTGTGGTC AGAAGAACTC TGGTGTTTGG	1727
TTCCCTGCTC AGCTGCGCGT GGACTGGGCT GAGCTCCTCA CCATACACTA GTGCCGGCTT	1787
TTGTTTCCTG TAAACAGTGG TTGCATGTGT AGAGAAGTAA CAAGCGAGTA TTCAGATCAT	1847
ACGAGGAGGC GTTCCTCGGT GCATGACGGT CAGATGGCCA TTTATCAGCA TATTTATTTG	1907
TATTTTCTCA GCACATAGTA AGGTACAAC TGTGTTTCTC AATTGTCTCG AAAAAACAGA	1967
GTTCTTAAGT GGCCCAGTTG TGGAGCCAAG TCTAAGTCGT GTGGAGTCAG TGCTGACATC	2027
ACTGGCTTGT GCTGTCTGTC ACATGTGTTT GTCTCTGCTG CTTGACCTCA TGGGATGTAC	2087
CCTCCAGTTC AACTGCCCCA AACAGACAGC CCCTTCCAAG CACCGTTCTT TGACAGCGGT	2147
AGCAGCTACC TATTCAAGAC GCCTCACACA AAATCTGCCT TAGAAAGTTA ATATATTTTA	2207
AATTATTTTA AAAGAACTC AACATCTTAT TCTTTGGCCT TTCTTAATG ATGCTTTATG	2267
GAGGCAGTGT TAACATTGTA CAGTGTATGC ATAGAGGAGT CTCCTCTATT TGAAGAACAA	2327
TGCAAAATGA GGCTTTCATT GAAGGGAAAA AAAAAAAAAA AA	2369

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

SUBSTITUTE SHEET (RULE 26)

- 139 -

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```

Met Glu Ala Gly Glu Glu Pro Leu Leu Leu Ala Glu Leu Lys Pro Gly
 1             5             10             15

Arg Pro His Gln Phe Asp Trp Lys Ser Ser Cys Glu Thr Trp Ser Val
          20             25             30

Ala Phe Ser Pro Asp Gly Ser Trp Phe Ala Trp Ser Gln Gly His Cys
          35             40             45

Val Val Lys Leu Val Pro Trp Pro Leu Glu Glu Gln Phe Ile Pro Lys
          50             55             60

Gly Phe Glu Ala Lys Ser Arg Ser Ser Lys Asn Asp Pro Lys Gly Arg
          65             70             75             80

Gly Ser Leu Lys Glu Lys Thr Leu Asp Cys Gly Gln Ile Val Trp Gly
          85             90             95

Leu Ala Phe Ser Pro Trp Pro Ser Pro Pro Ser Arg Lys Leu Trp Ala
          100             105             110

Arg His His Pro Gln Ala Pro Asp Val Ser Cys Leu Ile Leu Ala Thr
          115             120             125

Gly Leu Asn Asp Gly Gln Ile Lys Ile Trp Glu Val Gln Thr Gly Leu
          130             135             140

Leu Leu Leu Asn Leu Ser Gly His Gln Asp Val Val Arg Asp Leu Ser
          145             150             155             160

Phe Thr Pro Ser Gly Ser Leu Ile Leu Val Ser Ala Ser Arg Asp Lys
          165             170             175

Thr Leu Arg Ile Trp Asp Leu Asn Lys His Gly Lys Gln Ile Gln Val
          180             185             190

Leu Ser Gly His Leu Gln Trp Val Tyr Cys Cys Ser Ile Ser Pro Asp
          195             200             205

Cys Ser Met Leu Cys Ser Ala Ala Gly Glu Lys Ser Val Phe Leu Trp
          210             215             220

```

SUBSTITUTE SHEET (RULE 26)

- 140 -

Ser Met Arg Ser Tyr Thr Leu Ile Arg Lys Leu Glu Gly His Gln Ser
 225 230 235 240
 Ser Val Val Ser Cys Asp Phe Ser Pro Asp Ser Ala Leu Leu Val Thr
 245 250 255
 Ala Ser Tyr Asp Thr Ser Val Ile Met Trp Asp Pro Tyr Thr Gly Ala
 260 265 270
 Arg Leu Arg Ser Leu His His Thr Gln Leu Glu Pro Thr Met Asp Asp
 275 280 285
 Ser Asp Val His Met Ser Ser Leu Arg Ser Val Cys Phe Ser Pro Glu
 290 295 300
 Gly Leu Tyr Leu Ala Thr Val Ala Asp Asp Arg Leu Leu Arg Ile Trp
 305 310 315 320
 Ala Leu Glu Leu Lys Ala Pro Val Ala Phe Ala Pro Met Thr Asn Gly
 325 330 335
 Leu Cys Cys Thr Phe Phe Pro His Gly Gly Ile Ile Ala Thr Gly Thr
 340 345 350
 Arg Asp Gly His Val Gln Phe Trp Thr Ala Pro Arg Val Leu Ser Ser
 355 360 365
 Leu Lys His Leu Cys Arg Lys Ala Leu Arg Ser Phe Leu Thr Thr Tyr
 370 375 380
 Gln Val Leu Ala Leu Pro Ile Pro Lys Lys Met Lys Glu Phe Leu Thr
 385 390 395 400
 Tyr Arg Thr Phe

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1246 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

SUBSTITUTE SHEET (RULE 26)

- 141 -

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GACACTGCAT CGTCAAAC TG ATCCCCTGGC CGTTGGAGGA GCAGTTCATC CCTAAAGGGT	60
TTGAAGCCAA AAGCCGAAGT AGCAAAAATG AGACGAAAGG GCGGGGCAGC CCAAAAGAGA	120
AGACGCTGGA CTGTGGTCAG ATTGTCTGGG GGCTGGCCTT CAGCCTGTGC TTTCCCCACC	180
CAGCAGGAAG CTCTGGGCAC GCCACCACCC CCAAGTGCCC GATGTCTCTT GCC'TGGT'CT	240
TGCTACGGGA CTCAACGATG GGCAGATCAA GATCTGGGAG GTGCAGACAG GGCTCCTGCT	300
TTTGAATCTT TCCGGCCACC AAGATGTCGT GAGAGATCTG AGCTTCACAC CCAGTGGCAG	360
TTTGATTTTG GTCTCCGCGT CACGGGATAA GACTCTTCGC ATCTGGGACC TGAATAAACA	420
CGGTAAACAG ATTCAAGTGT TATCGGGCCA CCTGCAGTGG GTTACTGCT GTTCCATCTC	480
CCCAGACTGC AGCATGCTGT GCTCTGCAGC TGGAGAGAAG TCGGTCTTTC TATGGAGCAT	540
GAGGTCCTAC ACGTTAATTC GGAAGCTAGA GGGCCATCAA AGCAGTGTTG TCTCTTGTA	600
CTTCTCCCC GACTCTGCCC TGCTTGTCAC GGCTTCTTAC GATACCAATG TGATTATGTG	660
GGACCCCTAC ACCGGCGAAA GGCTGAGGTC ACTCCACCAC ACCCAGGTTG ACCCCGCCAT	720
GGATGACAGT GACGTCCACA TTAGCTCACT GAGATCTGTG TGCTTCTCTC CAGAAGGCTT	780
GTACCTTGCC ACGGTGGCAG ATGACAGACT CCTCAGGATC TGGGCCCTGG AACTGAAAAC	840
TCCCATTGCA TTTGCTCCTA TGACCAATGG GCTTTGCTGG CACATTTTTT CCACATGGTG	900
GAGTCATTGC CACAGGGACA AGAGATGGCC ACGTCCAGTT CTGGACAGCT CCTAGGGTCC	960
TGTCCTCACT GAAGCACTTA TGCCGAAAAG CCCTTCGAAG TTTCTAACA ACTTACCAAG	1020
TCCTAGCACT GCCAATCCCC AAGAAAATGA AAGAGTTCCT CACATACAGG ACTTTTAAAG	1080
CAACACCACA TCTTGTGCTT CTTTGTAGCA GGGTAAATCG TCCTGTCAA GGGAGTTGCT	1140
GGAATAATGG GCCA AACATC TGGTCTTGCA TTGAAATAGC ATTTCTTTGG GATTGTGAAT	1200

- 142 -

AGAATGTAGC AAAACCAGAT TCCAGTGTAC TAGTCATGGA TTTTTC

1246

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 422 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

ACCATGGTTC CAAGTCCTCT CCCCTGTGGT CAAGTTGCCC GAATGTTGGG CCCAAGTGCC	60
TTTTCTCCT TGGGCCTCCC CTTCTGACCT GCAGGACAGT TTTCCGGAGC CCATTTGGTA	120
TGAGGTATTA ATTAGCCTTA ACTAAATTAC AGGGGACTCA GAGGCCGTGC TCCTGACCGA	180
TCCAGACACT ATTTTTTTTT TTTTTTTTTA ACAATGGTGT GCATGTGCAG GAAATGACAA	240
ATTTGTATGT CAGATTATAC AAGGATGTAT TCTTAAACCG CATGACTATT CAGATGGCTA	300
CTGAGTTATC AGTGGCCATT TATTAGCATC ATATTTATTT GTATTTTCTC AACAGATGTT	360
AAGGTACAAC TGTGTTTTTC TCGATTATCT AAAAACCATA GTACTTAAAT TGAAAAAAAA	420
AA	422

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2019 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

SUBSTITUTE SHEET (RULE 26)

- 143 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGCACGAGGC GGGGTCAGGG CGGAGGCTGA GGACCAAGTA GGCATGGCGG AGGGCGGGAC	60
CGGCCCCGAT GGACGGGCCG GCCCGGGACC CGCAGGTCCT AATCTGAAGG AGTGGCTGAG	120
GGAGCAGTTC TGTGACCATC CACTGGAGCA CTGTGACGAT ACAAGACTCC ATGATGCAGC	180
CTATGTAGGG GACCTCCAGA CCCTCAGGAA CCTACTGCAA GAGGAGAGCT ACCGGAGCCG	240
CATCAATGAG AAGTCTGTCT GGTGCTGCGG CTGGCTTCCC TGCACACCAC TGAGGATCGC	300
AGCCACTGCA GGCCATGGGA ACTGTGTGGA CTTCTCATA CGCAAAGGGG CCGAGGTGGA	360
CCTGGTGGAT GTCAAGGGGC AGACTGCCCT GTATGTGGCT GTAGTGAACG GGCAC TTGGA	420
GAGCACTGAG ATCCTTTTGG AAGCTGGTGC TGATCCCAAC GGCAGCCGGC ACCACCGCAG	480
CACTCCTGTG TACCATGCCT YTCGTGTGGG TAGGGACGAC ATCCTGAAGG CTCTTATCAG	540
GTATGGGGCA GATGTTGATG TCAACCATCA TCTGAATTCT GACACCCGGC CCCCTTTTTTC	600
ACGGCGGCTA ACCTCCTTGG TGGTCTGTCC TCTATACATC AGTGCTGCCT ACCATAACCT	660
TCAGTGCTTC AGGCTGTCTT TGCAGGCTGG GGCAAATCCT GACTTCAATT GCAATGGCCC	720
TGTCAACACC CAGGAGTTCT ACAGGGGATC CCCTGGGTGT GTCATGGATG CTGTCTGCG	780
CCATGGCTGT GAAGCAGCCT TCGTGAGTCT GTTGGTAGAG TTTGGAGCCA ACCTGAACCT	840
GGTGAAGTGG GAATCCCTGG GCCCAGAGGC AAGAGGCAGA AGAAAGATGG ATCCTGAGGC	900
CTTGCAGGTC TTTAAAGAGG CCAGAAGTAT TCCCAGGACC TTGCTGAGTT TGTGCCGGGT	960
GGCTGTGAGA AGAGCTCTTG GCAAATACCG ACTGCATCTG GTTCCCTCGC TGCCGCTGCC	1020
AGACCCCAT AAGAAGTTT TGCTTTATGA GTAGCATTCA CATGCAGTGC TGA CTGCAAT	1080
GTGGAAGCCG ATCACCTGCA GTGAAACTG ACACAGACTC TGGCATCCTG GGAACCATGG	1140
CCTGTGCTGC CAGCTTGATC CTTGGCTGTC AGTGAAGAAA AAACGGCTGT GTTCTCTTGG	1200
ACTGTGATTC TATCTCAGGT GCTTGGGCCA TCGAACGCTC CTTGAGTCAT TGTCAACTGA	1260

SUBSTITUTE SHEET (RULE 26)

- 144 -

GAGGCACATA CAAACTTAAT TTTGTTCCCTC TTCAGTCTCT CTGTTTTGGA TTCTTCCTGG	1320
CAATGTGTGC AGCATGGGCT GAGCCTGGTG ATTGCCCTAG TGGGAAGGC TTTTCTCTCC	1380
AGGCTATGCA TCTATTTATG TTCCTACTTT GCAATTTATT GTTCTTTTAA GGCTTGATAT	1440
CAAAACAGAA AGAGGTTTGT TAAGAAAAGA TATAGGGAGA AAGGAATTCC GGTTCCTGTC	1500
ACTTGCTAGC CTGCTTTCCT TGCCTGGGTT TGTCTGTCTA TGCTGCCTGG TGCACATCCC	1560
TTCTCTTGC TGCCACTGTT CTATTTTGGG AGTTGTCTTC CGTCTAAGAT GGCTTCTGGG	1620
GTTCTATCTT ATTGCACAGA GGTCCCAGAA CAGTGTTTCAT AGGGCACCAT CTGCTCTGCC	1680
AAGGGTTTTC TGATGTCTTA CCCTGGGGAT CTCAGACAG TGGTTACCTT TAGGAGACCC	1740
ACCTGGAAC TACCATTAAAG TGAAGTCCCA CATTGAGATC AGGGACCATC TTAATAGTAC	1800
TCACTGCCAG TCCTCACAAG AGAAGATGAC ACGGGTGCTC TCTTCAGACA CTCCCATACA	1860
GGAAGTTGGA AAATGTCTTG GTCACCTGGG TTGTTCCCAG GCTACAACTT CTTGGTGTTC	1920
CACTAARACC AGRATATCCT AGTTTTTTGG GTTGAAGTGT CCCTCCCCAC TTTCTTGAA	1980
NCCCAATGCC CNTTTGKTN GGTGCTTCC CTAAAKTT	2019

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 350 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Ala	Arg	Gly	Gly	Val	Arg	Ala	Glu	Ala	Glu	Asp	Gln	Val	Gly	Met	Ala
1				5						10				15	

- 145 -

Glu Gly Gly Thr Gly Pro Asp Gly Arg Ala Gly Pro Gly Pro Ala Gly
 20 25 30

Pro Asn Leu Lys Glu Trp Leu Arg Glu Gln Phe Cys Asp His Pro Leu
 35 40 45

Glu His Cys Asp Asp Thr Arg Leu His Asp Ala Ala Tyr Val Gly Asp
 50 55 60

Leu Gln Thr Leu Arg Asn Leu Leu Gln Glu Glu Ser Tyr Arg Ser Arg
 65 70 75 80

Ile Asn Glu Lys Ser Val Trp Cys Cys Gly Trp Leu Pro Cys Thr Pro
 85 90 95

Leu Arg Ile Ala Ala Thr Ala Gly His Gly Asn Cys Val Asp Phe Leu
 100 105 110

Ile Arg Lys Gly Ala Glu Val Asp Leu Val Asp Val Lys Gly Gln Thr
 115 120 125

Ala Leu Tyr Val Ala Val Val Asn Gly His Leu Glu Ser Thr Glu Ile
 130 135 140

Leu Leu Glu Ala Gly Ala Asp Pro Asn Gly Ser Arg His His Arg Ser
 145 150 155 160

Thr Pro Val Tyr His Ala Xaa Arg Val Gly Arg Asp Asp Ile Leu Lys
 165 170 175

Ala Leu Ile Arg Tyr Gly Ala Asp Val Asp Val Asn His His Leu Asn
 180 185 190

Ser Asp Thr Arg Pro Pro Phe Ser Arg Arg Leu Thr Ser Leu Val Val
 195 200 205

Cys Pro Leu Tyr Ile Ser Ala Ala Tyr His Asn Leu Gln Cys Phe Arg
 210 215 220

Leu Leu Leu Gln Ala Gly Ala Asn Pro Asp Phe Asn Cys Asn Gly Pro
 225 230 235 240

Val Asn Thr Gln Glu Phe Tyr Arg Gly Ser Pro Gly Cys Val Met Asp
 245 250 255

SUBSTITUTE SHEET (RULE 26)

- 146 -

Ala Val Leu Arg His Gly Cys Glu Ala Ala Phe Val Ser Leu Leu Val
 260 265 270

Glu Phe Gly Ala Asn Leu Asn Leu Val Lys Trp Glu Ser Leu Gly Pro
 275 280 285

Glu Ala Arg Gly Arg Arg Lys Met Asp Pro Glu Ala Leu Gln Val Phe
 290 295 300

Lys Glu Ala Arg Ser Ile Pro Arg Thr Leu Leu Ser Leu Cys Arg Val
 305 310 315 320

Ala Val Arg Arg Ala Leu Gly Lys Tyr Arg Leu His Leu Val Pro Ser
 325 330 335

Leu Pro Leu Pro Asp Pro Ile Lys Lys Phe Leu Leu Tyr Glu
 340 345 350

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 419 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCATCCATGG CGGAGGGCGG CAGCACGACG GGCGGGCAGG GCCGGGCTCC GCAGGTCGTA	60
ATCTGAAGGA GTGGCTGAGG GAGCAATTTT GTGATCATCC GCTGGAGCAC TGTGAGGACA	120
CGAGGCTCCA TGATGCAGCT TACGTCGGGG ACCTCCAGAC CCTCAGGAGC CTATTGCAAG	180
AGGAGAGCTA CCGGAGCCGC ATCAACGAGA AGTCTGTCTG GTGCTGTGGC TGGCTCCCCT	240
GCACACCGTT GCGAATCGCG GCCACTGCAG GCCATGGGAG CTGTGTGGAC TTCCTCATCC	300
GGAAGGGGGC CGAGGTGGAT CTGGTGGACG TAAAAGGACA GACGGCCCTG TATGTGGCTG	360

SUBSTITUTE SHEET (RULE 26)

- 147 -

TGGTGAACGG GCACCTAGAG AGTACCCAGA TCCTTCTCGA AGCTGGCGCG GACCCCAAC 419

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 595 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GAGGAAGAAG AAAAGTGGAC CCTGAGGCCT TGCAGGTCTT TAAAGAGGCC AGAAGTGTTC	60
CCAGAACCTT GCTGTGTCTG TGCCGTGTGG CTGTGAGAAG AGCTCTTGGC AAAACCGGCT	120
TCATCTGATT CCTTCGCTGC CTCTGCCAGA CCCCATAAAG AAGTTTCTAC TCCATGAGTA	180
GACTCCAAGT GCTGCGGTTG ATTCCAGTGA GGGAGAAAGT GATCTGCAGG GAGGTGGACA	240
CCGAGCCCTG AGTGCTGTGC TGCTGCTGGT CTCCTGATGG CTGTTGCTGC AGAAGATGTC	300
CTCGTAGACT GTCATTGCTC CTCAGGTGCC TGGGCCGCTG AACAGTCCTT GGGTCATTGT	360
CAGCTGAGAG GCTTATACTA AAGTTATTAT TGTTTTTCCC AAGTTCTCTG TTCTGGATTT	420
TCAGTTGCAT ATTAATGTAA CGGGCCATGG GGTATGTACA TGTAGGGGCT GAGGTGGAG	480
GCCTACTAAT TTCCTGTAGG GAAGACTCCC AGCACTTCTG GAACTGTGCT TCTCTTTATT	540
TTTCTACTTC TCAATTGAT GGTTCGATTA AAGCCTTCTA GTATCTCAAT GAAAA	595

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 896 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- 148 -

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 4..396

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CTG ATG TCC GCA ATT CTG AAG GTT GGA CAC CAC TGC TGG CTG CCT GTG	48
Met Ser Ala Ile Leu Lys Val Gly His His Cys Trp Leu Pro Val	
1 5 10 15	
ACA TCC GCT GTC AAT CCC CAA AGG ATG CTG AGG CCA CCA CCA ACC GCT	96
Thr Ser Ala Val Asn Pro Gln Arg Met Leu Arg Pro Pro Pro Thr Ala	
20 25 30	
GTT TTC AAC TGT GCC GCT TGC TGC TGT CTG TGG GGG CAG ATG CTG ATG	144
Val Phe Asn Cys Ala Ala Cys Cys Cys Leu Trp Gly Gln Met Leu Met	
35 40 45	
AAT ACA TAC CGT GTA GTT CAG CTT CCT GAG GAG GCC AAG GGC TTG GTG	192
Asn Thr Tyr Arg Val Val Gln Leu Pro Glu Glu Ala Lys Gly Leu Val	
50 55 60	
CCA CCA GAG ATT CTA CAG AAG TAC CAT GGA TTC TAC TCT TCC CTC TTT	240
Pro Pro Glu Ile Leu Gln Lys Tyr His Gly Phe Tyr Ser Ser Leu Phe	
65 70 75	
GCC TTG GTG AGG CAG CCC AGG TCG CTG CAG CAT CTC TGC CGT TGT GCG	288
Ala Leu Val Arg Gln Pro Arg Ser Leu Gln His Leu Cys Arg Cys Ala	
80 85 90 95	
CTC CGC AGT CAC CTG GAG GGC TGT CTG CCC CAT GCA CTA CCG CGC CTT	336
Leu Arg Ser His Leu Glu Gly Cys Leu Pro His Ala Leu Pro Arg Leu	
100 105 110	
CCC CTG CCA CCG CGC ATG CTC CGC TTT CTG CAG CTG GAC TTT GAG GAT	384
Pro Leu Pro Pro Arg Met Leu Arg Phe Leu Gln Leu Asp Phe Glu Asp	
115 120 125	
CTG CTC TAC TAGGCTTGCT GCCCTGTGAA CAAAGCAGAC CCCACCCCCA	433
Leu Leu Tyr	
130	

SUBSTITUTE SHEET (RULE 26)

- 149 -

```

CCCCAAGGGC ATCTCTCAGC AATGAATGAT GCAAGGCGGT CTGTCTTCAA GTCAGGAGTG      493
GACGCCTTGA TCCACACTTG AGAGAAGAGG CCAGATCAGC ACCYGGCTGG TAGTGATNGC      553
AGAGGGCACC TGTGCAGATC TGTGTGCGCA CTGGAATCT CTAGGCTGAA GGCYAGAGCA      613
AATGGTGCAR GTGTTAGTCC TTGGGANGAG AGACAGANGG TGAGAAAGCA AGACAGAGGT      673
GAGAGTGCAC ATGTCAAGTG GTAGATTGCC TTAAGAGAAA GCTAAAAAAA GAAAAAGATT      733
CGGGCGAACT TCTTTAGGGG TAATGCTGCA GCGTGTTAAA CTGACTGACC AGCGTCCATA      793
TCTTTGGACC CTTCCCGGGT GAAAAAGCCC CTTATCCTC CAGCGCTCCC CAAGGGTGCT      853
TAGCAATACC GGGTGCTTTT CTGCCGCAA GTGAGTTACC AAA                          896

```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 130 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```

Met Ser Ala Ile Leu Lys Val Gly His His Cys Trp Leu Pro Val Thr
 1           5           10           15
Ser Ala Val Asn Pro Gln Arg Met Leu Arg Pro Pro Pro Thr Ala Val
          20           25           30
Phe Asn Cys Ala Ala Cys Cys Cys Leu Trp Gly Gln Met Leu Met Asn
          35           40           45
Thr Tyr Arg Val Val Gln Leu Pro Glu Glu Ala Lys Gly Leu Val Pro
          50           55           60
Pro Glu Ile Leu Gln Lys Tyr His Gly Phe Tyr Ser Ser Leu Phe Ala
          65           70           75           80
Leu Val Arg Gln Pro Arg Ser Leu Gln His Leu Cys Arg Cys Ala Leu
          85           90           95

```

SUBSTITUTE SHEET (RULE 26)

- 150 -

Arg Ser His Leu Glu Gly Cys Leu Pro His Ala Leu Pro Arg Leu Pro
 100 105 110

Leu Pro Pro Arg Met Leu Arg Phe Leu Gln Leu Asp Phe Glu Asp Leu
 115 120 125

Leu Tyr
 130

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 436 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTGGGGGCGT CATCATGACC TCCTCTAGGG CTCTGCAACA TGACTCCTGT GGTGCAAATC	60
AACAAATTGT TCACTGATGA ATCCACAAGG ATCTCTGGGC CTACAACCAG GTCCTGGTCC	120
ACATGACTGT CGTCTTCGGA GAAGGCACCA CTCGCCCCCG GCAGGTACGG CTGACACCTC	180
CATGGGAGAA GACGTATCCA GGCAGCAGCT GCGCGGCCCT TCAAGAGGGC ACATCCCGTC	240
ATCTAAAGGC ACGGTGTACT GAAGGTAGTC CTGAGACATG AGTCCGATTA CTACAGGCAC	300
GTGTTCTCC AGGTGGAGGC TCAGGTCCCC GGGTGAGCTG GGGCTGCAGC GGGACTCAGG	360
GCGCGGCTCT GGCTGCAGGT CTCGCAGCTC CCTGGGCTGT AGCTCCCGCA GATCCTTGCG	420
CACACCGTTG ACTGGT	436

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- 151 -

- (A) LENGTH: 2180 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TTAATAGTAC CTACATAGTA GAAAATTATA ACTCCACTTT AAAACAATGT TTTCTTTCTA	60
TTCAAATCAA TTTAAACTT TTTATAAACA TTAATGTTGC AAGAGAATCC AGTCCATTTA	120
TGAAAATTAG TTGACAATCA AGTTCACCCA AGAAAATGTT GACTAAGCTA AAGAAATCAC	180
AGATAAAACA TTTTACCAA AGGATAGGTA ACACACAAAA AAATGCTATC ACAGGAAGCT	240
ATGATCATCT AATATTTCTT TAATAATAAT TCTAGTTCCA TAGGTTTTC TGTATGCCA	300
ATTTGTACCC GAGTTTAATT ACAGAAAAGG CAACAATTC TAAATTGGTG GTATACATT	360
CTTTACAATT TTTTAATGTA AGGCCATTTA TTTAAATAGA CAACTAGAA GATGAAAACG	420
AAGGCAACAG AAAAATTCAA CTTTTCACAA CAAAAGAAT TAGCACAACC TTAGAAATAA	480
TTTAGAAAAA AGTGTGTGTA AAAGATATGT TGCAGATCTC CGTTCATTA CCCAAGATTA	540
TGTCAATTCA CGATTCTAAA TAAATCTTTT TAAAGTAAGA GATTAAAAAC TCATCTTCAG	600
TGTATATGTA AATCCGTGG TTTTATCACA CAGGTATGTT TATTCAACAC TGCTTTGGAA	660
ATGGACCATT TAAAAGGACA TGGCAATTTC CATTCTGTTA AGTTTCATTC AACCTTTACT	720
TAGGGGTTGA TTACCACATG AAATGTGCTT TTAATGCATA AAAATCACAG TGGATTAGCC	780
AGCAAAAGGG ACTGGGCGGG GGGGCGATTG AGGAGAATTT GATAATTCAC ATTGTGATTA	840
TTCTGCACAT TGATGAAACA TAATTCACAC CTCTAAAACC TCAAGACTTC CCTTTTAA	900
AGAACCACAAA TAAACCCAAG ACACCTTGCT GACACTTCCC CACCCCTAAA CAACTGATG	960
ACTCTTTTAC ACATAAACT GAAATAGTTA TGGCAGCAA AGATTTTGAT GGCAATGAAA	1020

- 152 -

GTTTGTAAC TGTATTTCAA TCTCTGTTC TTATTCCTAA AGTGCAAGAT GCAGGGTTCT	1080
CAATCTTTCA GTAGTGCTTC TCCTGTAAAT AATCCTTCAT TTTGTTTGGC AAAGGCAGTT	1140
TCTGAATTAA GTCTATTCTG GTATACTGAC GTATAACAAA ACGACACAGG TACTGCAACG	1200
AGCGCACCTA TGAACCCCGG AACACTGGTT GGCAAGTTCT GACGGAAGTG CAGATTCCAG	1260
GCAGCGAGAC CTTGAATAAC AAAAAGCTCC CATTTTCAGA GTCCCTGATT GAATGCTCCA	1320
ATTAGATCAA CTATGGACGT ATGTCCTTCC ACATCGGCTG TTCATAAAAG CTAAACCTAC	1380
CATTTGAGTG CTCAATTCTA GTGTGAAGTG TTTTACCATG GGAGCGAAAG TCACAGCTTA	1440
AAAGGTAACG GTCGTCAGAA CTGTCCCGAA CAAGAAAAGA ACCATCTGGC ACGTTTGCTA	1500
GCTTCCCTTC TGCCTCCCAA CGTGTGATTG GTCCCCAGTA CCATCCTTGC TTTGCAAGTT	1560
TTTTCAGCTC CTCTGTAAGG CTTGTCACAA CCATGGGACC ACTACTTTGC ACTGAGTCAT	1620
AAACTCTTGC AACCCAGGA GCAGAGTTCG GATCAAAATT CAAATGACAG CGCATAACTT	1680
TCAGCCACGT GGGGCTTTCT GTCCAGTGAG TCCACTGAAA GTTCCCCTTT GGGATTTGGA	1740
TTATTCCTGC ATTGGAGTAA CCAATGGTGA AGATTGGAGG GACATCCATC GTGAACCCGC	1800
TCTCCGGGGT TCTGCAACAT GACTCCCGTG GTGCCAATCA ACAAGCCATT CACCGGACTG	1860
ATCCACGAAG ATCTCTGGGG CGACAAC TAGTCTGGTCT ACCTGACTCT CATCCTCGGG	1920
GAAAGCGCGC CCTCCCACTT GAGGAGGAAC CGCAGAGACT TCCATGGGAG AAGAGCTGTC	1980
CAGACAATAG CTCCGTGATC CTTCCAAAGG ATACATCCCC TCATCTAAAG GCACAGTATA	2040
CTGAATGTAG TCCTGAGGCA TAAGTCCAAT AACGACAGGC ACATGTTCAT CCAGGTGAAG	2100
ATGCAGGTCT CCATTATGAG AAGCCGAGCT CTTCAAGTAA TTGGCTTGCT CCTGGCACGT	2160
GGTCTCAGAC TGGAGGTCGT	2180

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

SUBSTITUTE SHEET (RULE 26)

- 153 -

- (A) LENGTH: 2649 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```

GGCACGAGGC TGTGTCCAGC ACACAGAGAG GGCCCGGCCA TCTGCTTTGG TTCAGAGCCC      60
TGTGTCTGTC TGTCACCTAG ACTCTTCCTC CCGGCTCGCA GCTCACCTC CATCCTCCTT      120
ACTGGCTCCA GCATGACTCG CTTCTCTTAT GCAGAGTACT TTGCTCTGTT TCACTCTGGC      180
TCTGCACCTT CCAGGTCCCC TTCGTCTCCC GAGAACCCAC CGGCCCCGCGC ACCCCTGGGT      240
CTGTTCCAAG GGGTCATGCA GAAGTATAGC AGCAACCTGT TCAAGACCTC CCAGATGGCG      300
GCTATGGACC CCGTGCTGAA GGCCATCAAG GAAGGGGATG AAGAGGCCTT GAAGATCATG      360
ATCCAGGATG GGAAGAATCT TGCAGAGCCC AACAAGGAGG GCTGGCTGCC GCTCCACGAG      420
GCTGCCTACT ATGGCCAGCT GGGCTGCCTG AAAGTCCTGC AGCAAGCCTA CCCAGGGACC      480
ATTGACCAAC GCACACTGCA GGAAGAGACA GCATTATACC TGGCCACATG CAGAGAACAC      540
CTGGATTGCC TCCTGTCGCT GCTCCAGGCG GGGGCAGAGC CTGACATCTC TAACAAATCC      600
AGGGAGACTC CACTTTACAA AGCCTGTGAG CGCAAGAACG CGGAGGCGGT GAGGATATTG      660
GTGCGATACA ACGCAGACGC CAACCACCGC TGTAACAGGG GCTGGACCGC ACTGCACGAG      720
TCTGTCTCCC GCAATGACCT GGAGGTCATG GAGATCCTAG TGAGTGGCGG GGCCAAGGTG      780
GAGGCCAAGA ATGTCTACAG CATCACCCCT TTGTTTGTGG CTGCCCAGAG TGGGCAGCTG      840
GAGGCCCTGA GGTTCCTGGC CAAGCATGGT GCAGACATCA ACACGCAGGC CAGTGACAGT      900
GCATCAGCCC TCTACGAGGC CAGCAAGAAT GAGCATGAAG ACGTGGTAGA GTTTCTTCTC      960
TCTCAGGGCG CCGATGCTAA CAAAGCCAAC AAGGACGGCC TGCTCCCCCT GCATGTTGCC      1020

```


- 154 -

TCCAAGAAGG GCAACTATAG AATAGTGCAG ATGCTGCTGC CTGTGACCAG CCGCACGCGC	1080
GTGCGCCGTA GCGGCATCAG CCCGCTGCAC CTAGCGGCCG AGCGCAACCA CGACGCGGTG	1140
CTGGAGGCGC TGCTGGCCGC GCGCTTCGAC GTGAACGCAC CTCTGGCTCC CGAGCGCGCC	1200
CGCCTCTACG AGGACCGCCG CAGTTCTGCG CTCTACTTCG CTGTGGTCAA CAACAATGTG	1260
TACGCCACCG AGCTGTTGCT GCTGGCGGGC GCGGACCCCA ACCGCGATGT CATCAGCCCT	1320
CTGCTCGTGG CCATCCGCCA CGGCTGCCTG CGCACCATGC AGCTGCTGTT GGACCATGGC	1380
GCCAACATCG ACGCCTACAT CGCCACTCAC CCCACCGCCT TTCCAGCCAC CATCATGTTT	1440
GCCATGAAGT GCCTGTCGTT ACTCAAGTTC CTTATGGACC TCGGCTGCGA TGGCGAGCCC	1500
TGCTTCTCCT GCCTGTACGG CAACGGGCCG CACCACCCGC CCCGCGACCT GGCCGCTTCC	1560
ACGACGCACC CGTGGACGAC AAGGCACCTA GCGTGGTGCA GTTCTGTGAG TTCCTGTGCG	1620
CCCCGGAAGT GAGCCGCTGG GCGGGACCCA TCATCGATGT CCTCCTGGAC TATGTGGGCA	1680
ACGTGCAGCT GTGCTCCCGG CTGAAGGAGC ACATCGACAG CTTTGAGGAC TGGGCTGTCA	1740
TCAAGGAGAA GGCAGAACCT CCGAGACCTC TGGCTCACCT CTGCCGGCTG CGGGTTCGGA	1800
AGGCCATAGG AAAATACCGG ATAAACTCC TGGACACACT GCCGCTTCCC GGCAGGCTAA	1860
TCAGATACTT GAAATATGAG AATACACAGT AACCAGCCTG GAGAGGAGAT GTGGCCTTCA	1920
GA CTGTTTCC GGGACGCCCC AGGTGGCCTG CATCCAGGAC CCCCTGGGGT CAGAACAGGT	1980
GTGACCTTGC TGGTTCTTTG CTGGAGCTTC ACCCAAAGTG AGAACCTGAT GTGGGGAGTG	2040
GACGTGGAAC CTCTGCTTTC AACTGTGTCAG CGGATCGCAG ACCCGCTCTG CTCTTGCCA	2100
TAGCCAGAGA CCTTCAACCT GGGGCCAGGG GAGAGCTGGT CTGGGCAAGG TGGCCAGGC	2160
AGGAATCCTG GCCTTAAGCT GGAGAACTTG TAGGAATCCC TCACTGGACC CTCAGCTTTC	2220
AGGCTGCGAG GGAGACGCCC AGCCCAAGTA TTTTATTTCC GTGACACAAT AACGTTGTAT	2280
CAGAAAAAAA AAAAAACATG GGCGCAGCTT ATTCCTTAGT AGGGTATTTA CTTGCATGCG	2340
CGCTTAAAGC TACTGGAAAC ATGCGTTCCA CTATGCTTGA GAATCCCCTT GCACTGGTAA	2400

- 155 -

ACGAGAGCCG ACGTGCTTCA AGGTTGGATT TTTGGTTGCC CCTTGGCGT TCCGCGGGTT	2460
TGTCCGACGT AATTGACCCC GTGTTTTGTC ACTTTCGAGT GTTCCGACTA TTGGGGGGCT	2520
TTTGGTTGTC CCCAAAATTG TGGGTGGTGT GCGGACGCCA CGAGAAAGTG TTCATGGGCG	2580
ATAATCATT A CTGGAGAATG TAGAGCGGCG GTTTTACGAA TAAATATTTT TTAAGCCGCC	2640
TTCCCAAAA	2649

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 495 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CCTCCTGAGA GTTCGCCGGC CCGGGCCCAA TGGGTTGTTC CAAGGGGTCA TGCAGAAATA	60
CAGCAGCAGC TTGTTCAAGA CCTCCCAGCT GGCGCCTGCG GACCCCTTGA TAAAGGCCAT	120
CAAGGATGCG ATGAAGAGGC CTTGAAGACC ATGATCAAGG AAGGAAGAA TCTCGCAGAG	180
CCCAACAAGG AGGGCTGGCT GCCGCTGCAC GAGGCCGCAT ACTATGGCCA GGTGGGCTGC	240
CTGAAAGTCC TGCAGCGAGC GTACCCAGGG ACCATCGACC AGCGCACCCCT GCAGGAGGAA	300
ACAGCCGTTT ACTTGGCAAC GTGCAGGGGC CACCTGGACT GTCTCCTGTC ACTGCTCCAA	360
GCAGGGGCAG AGCGGGACAT CTCCAACAAA TCCCGAGAGA ACCGCTCTAC AAAGCCTGTG	420
AGCGCAAGAA CGCGGAAGCC GTGAAGATTC TTGGTGCAGC ACAACGCAGA CACCAACAAC	480
GCTGCAACCG GGCTG	495

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- 156 -

- (A) LENGTH: 709 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```

GTGCAGCTCT GCTCGCGGCT GAAGGAACAC ATCGACAGCT TTGAGGACTG GGCCGTCATC      60
AAGGAGAAGG CAGAACCTCC AAGACCTCTG GCTCACCTTT GCCGACTGCG GGTTCGAAAG      120
GCCATTGGGA AATACCGTAT AAAACTCCTA GACACCTTGC CGCTCCCAGG CAGGCTGATT      180
AGATACCTGA AATACGAGAA CACCCAGTAA CTGGGGCCAC GGGGAGAGAG GAGTAGCCCC      240
TCAGACTCTT CTTACTAAGT CTCAGGACGT CGGTGTTCCC AACTCCAAGG GGACCTGGTG      300
ACAGACGAGG CTGCAGGCTG CCTCCCTCTC AGCCTGGACA GCTACCAGGA TCTCACTGGG      360
TCTCAGGGCC CAGAGCTTTG GCCAGAGCAG AGAACAGAAT GTGTCAAGGA GAAGAATCAT      420
TTGTTTACAA ACTGATGAGC AGATCCCAGA CCTTCTCTAC CTTCAGGAAT GGCAGAAACC      480
TCTATTCCTG GGGCCAGGGC AGAGCTTGAG GTGTTCTGGG GAAGGTGGTG CTCAGAGCCT      540
TCCCTGTGCC CCTCCACTTG TTCTGGAAAA CTCACCACTT GACTTCAGAG CTTTCTCTCC      600
AAAGACTAAG ATGAAGACGT GGCCCAAGGT AGGGGGTAGG GGGAGCCTGG GTCTTGGAGG      660
GCTTTGTTAA GTATTAATAT AATAAATGTT ACACATGTGA AAAAAAAAAA      709

```

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 848 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- 157 -

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..624

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

TTG GAG AAG TGT GGT TGG TAT TGG GGG CCA ATG AAT TGG GAA GAT GCA	48
Leu Glu Lys Cys Gly Trp Tyr Trp Gly Pro Met Asn Trp Glu Asp Ala	
1 5 10 15	
GAG ATG AAG CTG AAA GGG AAA CCA GAT GGT TCT TTC CTG GTA CGA GAC	96
Glu Met Lys Leu Lys Gly Lys Pro Asp Gly Ser Phe Leu Val Arg Asp	
20 25 30	
AGT TCT GAT CCT CGT TAC ATC CTG AGC CTC AGT TTC CGA TCA CAG GGT	144
Ser Ser Asp Pro Arg Tyr Ile Leu Ser Leu Ser Phe Arg Ser Gln Gly	
35 40 45	
ATC ACC CAC CAC ACT AGA ATG GAG CAC TAC AGA GGA ACC TTC AGC CTG	192
Ile Thr His His Thr Arg Met Glu His Tyr Arg Gly Thr Phe Ser Leu	
50 55 60	
TGG TGT CAT CCC AAG TTT GAG GAC CGC TGT CAA TCT GTT GTA GAG TTT	240
Trp Cys His Pro Lys Phe Glu Asp Arg Cys Gln Ser Val Val Glu Phe	
65 70 75 80	
ATT AAG AGA GCC ATT ATG CAC TCC AAG AAT GGA AAG TTT CTC TAT TTC	288
Ile Lys Arg Ala Ile Met His Ser Lys Asn Gly Lys Phe Leu Tyr Phe	
85 90 95	
TTA AGA TCC AGG GTT CCA GGA CTG CCA CCA ACT CCT GTC CAG CTG CTC	336
Leu Arg Ser Arg Val Pro Gly Leu Pro Pro Thr Pro Val Gln Leu Leu	
100 105 110	
TAT CCA GTG TCC CGA TTC AGC AAT GTC AAA TCC CTC CAG CAC CTT TGC	384
Tyr Pro Val Ser Arg Phe Ser Asn Val Lys Ser Leu Gln His Leu Cys	
115 120 125	
AGA TTC CGG ATA CGA CAG CTC GTC AGG ATA GAT CAC ATC CCA GAT CTC	432
Arg Phe Arg Ile Arg Gln Leu Val Arg Ile Asp His Ile Pro Asp Leu	
130 135 140	

- 158 -

[illegible]

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 207 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Leu	Glu	Lys	Cys	Gly	Trp	Tyr	Trp	Gly	Pro	Met	Asn	Trp	Glu	Asp	Ala
1				5					10					15	
Glu	Met	Lys	Leu	Lys	Gly	Lys	Pro	Asp	Gly	Ser	Phe	Leu	Val	Arg	Asp
			20					25					30		
Ser	Ser	Asp	Pro	Arg	Tyr	Ile	Leu	Ser	Leu	Ser	Phe	Arg	Ser	Gln	Gly
		35					40					45			

- 159 -

Ile	Thr	His	His	Thr	Arg	Met	Glu	His	Tyr	Arg	Gly	Thr	Phe	Ser	Leu	50	55	60	
Trp	Cys	His	Pro	Lys	Phe	Glu	Asp	Arg	Cys	Gln	Ser	Val	Val	Glu	Phe	65	70	75	80
Ile	Lys	Arg	Ala	Ile	Met	His	Ser	Lys	Asn	Gly	Lys	Phe	Leu	Tyr	Phe	85	90	95	
Leu	Arg	Ser	Arg	Val	Pro	Gly	Leu	Pro	Pro	Thr	Pro	Val	Gln	Leu	Leu	100	105	110	
Tyr	Pro	Val	Ser	Arg	Phe	Ser	Asn	Val	Lys	Ser	Leu	Gln	His	Leu	Cys	115	120	125	
Arg	Phe	Arg	Ile	Arg	Gln	Leu	Val	Arg	Ile	Asp	His	Ile	Pro	Asp	Leu	130	135	140	
Pro	Leu	Pro	Lys	Pro	Leu	Ile	Ser	Tyr	Ile	Arg	Lys	Phe	Tyr	Tyr	Tyr	145	150	155	160
Asp	Pro	Gln	Glu	Glu	Val	Tyr	Leu	Ser	Leu	Lys	Glu	Ala	Gln	Arg	Gln	165	170	175	
Phe	Pro	Asn	Arg	Ser	Lys	Arg	Trp	Asn	Pro	Pro	Arg	Ser	Glu	Gly	Leu	180	185	190	
Pro	Ala	Gly	His	His	Gln	Gly	His	Leu	Val	Ala	Lys	Leu	Gln	Leu		195	200	205	

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 464 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

- 160 -

```

GTTCCAAGCC TAACCCATCT TTGTCGTTTG GAAATTCGGG CCAGTCTAAA AGCAGAGCAC      60
CTTCACTCTG ACATTTTCAT CCATCAGTTG CCACTTCCCA GAAGTCTGCA GAACTATTTG     120
CTCTATGAAG AGGTTTAAAG AATGAATGAG ATTCTAGAAC CAGCAGCTAA TCAGGATGGA      180
GAAACCAGCA AGGCCACCTG ACACAGGTCC TTTAATTCTG TTTAGTCACA AAAGACGGCT      240
TGTGTGACTG TTTGGATTTG GTGATCAAAT GTCCATGTTT ACAGTTGCTT TTCCCAGTTT      300
GTGTCTTTCC CAATATTGTG AACCTTATCC ATCTTGCCCTT ACTCAGTTTT ATTTCTAGTG      360
CACTTTGTTG TGTATTATTT GTTTACCTGA CCATTTTCTA CTTTATTCTG CTAATAAACT      420
GTAATTCTGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA                        464

```

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 747 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

```

GGGGATCGAA AGCGGGGGCT TCTGGGACGC AGCTCTGGAG ACGCGGCCTC GGACCAGCCA      60
TTTCGGTGTA GAAGTGGCAG CACGGCAGAC TGGTCAAACA AATGGATTTT ACAGAGGCTT     120
ACGCGGACAC GTGCTCTACA GTTGGACTTG CTGCCAGGGA AGGCAATGTT AAAGTCTTAA      180
GGAAACTGCT CAAAAGGGC CGAAGTGTCG ATGTTGCTGA TAACAGGGGA TGGATGCCAA      240
TTCATGAAGC AGCTTATCAC AACTCTGTAG AATGTTTGCA AATGTTAATT AATGCAGATT      300
CATCTGAAAA CTACATTAAG ATGAAGACCT TTGAAGGTTT CTGTGCTTTG CATCTCGCTG      360
CAAGTCAAGG ACATTGAAA ATCGTACAGA TTCTTTTAGA AGCTGGGGCA GATCCTAATG      420
CAACTACTTT AGAAGAAACG ACACCATGTG TTTTAGCTGT TGAAAATGGA CAGATAGATG      480

```

- 161 -

TGTTAAGGCT GTTGCTTCAA CACGGAGCAA ATGTTAATGG ATCCCATTCT ATGTGTGGAT	540
GGAACCTCCTT GCACCAGGCT TCTTTTCAGG AAAATGCTGA GATCATAAAA TTGCTTCTTA	600
GAAAAGGAGC AAACAAGGAA TGCCAGGATG ACTTTGGAAT CACACCTTTA TTTGTGGCTG	660
CTCAGTATGG CCAAGCTAGA AAGCTTTGAA GCATACTTAT TTCATCCGGG TGCAAATGTC	720
AATTGTCAAG CCTTGGACAA AGCTACC	747

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1018 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CACAAATGGG ACCATACAAA AATCTTGGAC TTGTTAATAA CCACTTACTA ACCGGGACCT	60
GTGACACTGG GCTAAACAAA GTAAGTCCCT GTTTACTCAG CAGTGTTTGG GGGACATGAA	120
GGATTGCCTA GAAATATTAC TCCGGAATGG TCTACAGCCC AGACGCCAG GCGTGCCTTG	180
TTTTTGATT CAGTTCTCCT GTGTGCATGG CTTTCCAAAA GGAGGTGGAG CTGTAGTTCT	240
TTGGAATTGT GAACATTCTT TTGAAATATG GAGCCCAGAT AAATGAACTT CATTTGGCAT	300
ACTGCCTGAA GTACGAGAAG TTTTCGATAT TTCGCTACTT TTTGAGGAAA GGTGCTCAT	360
TGGGACCATG GAACCATATA TATGAATTG TAAATCATGC AATTAAAGCA CAAGCAAAAT	420
ATAAGGAGTG GTTGCCACAT CTTCTGGTTG CTGGATTGA CCCACTGATT CTA CTGTGCA	480
ATTCTTGGAT TGA CT CAGTC AGCAT TGACA CCCTTATCTT CACTTTGGAG TTTACTAATT	540
GGAAGACACT TGCACCAGCT GTTGAAAGGA TGCTCTCTGC TCGTGCCTCA AACGCTTGGA	600
TTCTACAGCA ACATATTGCC CACTGTTCCA TCCCTGACCC ATCTTTGTGC TTTGGA AATT	660

- 162 -

CGGTCCAGTC TAAAATCAGA ACGTCTACGG TCTGACAGTT ATATTAGTCA GCTGCCACTT	720
CCCAGAAGCC TACATAATTA TTTGCTCTAT GAAGACGTTT TGAGGATGTA TGAAGTTCCA	780
GAAGTGGCAG CTATTCAAGA TGGATAAATC AGTGAAACTA CTTAACACAG CTAATTTTTT	840
TCTCTGAAAA ATCATCGAGA CAAAAGAGCC ACAGAGTACA AGTTTTTATG ATTTTATAGT	900
CAAAAGATGA TTATTGATTG TCAGATAGGT TAGGTTTTGG GGGGCCAGTA GTTCAGTGAG	960
AATGTTTATG TTTACAACTA GCCTTCCCAG TAAAAAAAAA AAAAAAAAAA AAAAAAAA	1018

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1897 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CGGGGGGCTG GGACCTGGGG CGTAACCGTC TCTACCACGA CGGCAAGAAC CAGCCAAGTA	60
AAACATACCC AGCCTTTCTG GAGCCGGACG AGACATTCAT TGTCCCTGAC TCCTTTTTCG	120
TGGCCCTGGA CATGRATGAT GGGACCTTAA GTTTCATCGT GGATGGACAG TACATGGGAG	180
TGGCTTTCCG GGGACTCAAG GGTAAAAAGC TGTATCCTGT AGTGAGTGCC GTCTGGGGCC	240
ACTGTGAGAT CCGCATGCGC TACTTGAACG GACTTGATCC TGAGCCCCTG CCACTCATGG	300
ACCTGTGCCG GCGTTCGGTG CGCCTAGCGC TGGGAAAAGA GCGCCTGGGT GCCATCCCCG	360
CTCTGCCGCT ACCTGCCTCC CTCAAAGCCT ACCTCCTCTA CCAGTGATCC ACATCCCAGG	420
ACCGCCATAC GACAGCCATC TGGTGCCAAR TCACTGAGCC CGTTGGGGTC CGCCGACCCC	480
TGCGCCTGGG ATGGAAGCCC ACCTCAGCCA TGGGCAGACG TGCCCCCTCA TCCTACCGGC	540

- 163 -

TGCCTCTGCT GGGGGAACCT ATGCCAACGG ACTTCTCCCT TCCCAACACT GGCTGAAGCA	600
GCAGCACCCA GGCCCTTCCC TGAACCAGAT GCAGAGAATA AACTATGAAA ACCTCTCTCA	660
GGCGCCTTCT GCTCTCAGGT GGAGTGGGCT GCCCCCACT CTCTGCAGAG AGAGGCTACA	720
CCCACCTGGG GGGTCCTGGG AGGTAAGACT AGTAGGAGGT GCCAGGGCTG ARTCCAAAAG	780
CAGGAATGGC CAGGAMCAGG CCATACAGAT GAAGCTCAGG ATGTCACATA CCATGGACAM	840
TGAGACAGAA CCCCAGGTTG GAMTTCCCTT GGGCCAACGA GTGCCAGCTT TAATGTCAGC	900
TGCMGGTGCT CTGTGGCCTG TATTTATTCT TTAAACAGTA GCAAAGGCCA TTTATTTATT	960
CCACTTAGAA AGGAAACCTT GGTGGGTGGY TTCCCTCGAT GTGCTTTCCC CCACCTCCCT	1020
GGAATGTGTG TGCCACACCT GTCCTTGTC CAGGCCAGGA CTGTGGCACA TGAGCTGGTG	1080
TGCACAGATA CACGTATGTC GTCGTGCATG ACCCCTGACT AGTTCCTAAG TAGCCCTGCA	1140
CCAAGCACCA GAGCAGACCC CAAGAGAGGC CCGTGCAAGT CCCCATGTCC CCAGGTCCCT	1200
GCTTCTGTTG CCTTGGGACT CATAACCCG CACACGTGTT TCAGCCTCTT GACTTCCATG	1260
AGCTTCGAAT TTTGCCCCCG ATTCTTCTGA TATTTCCCAT TGGCATCCTC CAAAGCTCTG	1320
GGCCTGGAGG GCATTAGGAC ACATGGAATG AGTGGGGTCT CCAGCCCCTG GGAAAGCCAC	1380
TGGCAAGGCA GGATTAGAAA GACCAAGAGC AGGGTGGGGC GCCATGAAGC CTGTATGCCT	1440
CTCAGGCTCA AGACCCCGCC ACACACCCAC TCAAGCCTCA GAAGTGGTGT GTAGGGCAGC	1500
CCCAGGAGAG GAATGCCTGT CCTAGCAGCA CGTACATGGA GCACCCACCA TGTGCTCCAG	1560
CCCTCTGGCT GTTCTCTTGT CTCTAGAATC AACTCCCTAC ATTGGGAATG TAGCCATTGT	1620
GTAGAGGACT TGCCTAGCCT GCAGGAAGCT CACGTTCCAT CCCCTGCACC AAGGAGAATC	1680
AAAGCTCAGG AGGCTGAGGC AGGAGGATTG CTGTCAAGTG TGTACAGAGG TCATGGCCAT	1740
CCTGGGCTAT ATTAACCTT GTCCTTTAAG AAAAAGAAAA GAAATCAACT TCCATTGAAT	1800
CTGAGTTCTG CTCATTTCTG CACAGGTACA ATAGATGACT TKATTTGTTG AAAAATGKTT	1860
AATATATTTA CMTATATATA TATTTGTAAG AAGCATT	1897

- 164 -

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

```

Gly Gly Trp Asp Leu Gly Arg Asn Arg Leu Tyr His Asp Gly Lys Asn
1           5           10           15
Gln Pro Ser Lys Thr Tyr Pro Ala Phe Leu Glu Pro Asp Glu Thr Phe
          20           25           30
Ile Val Pro Asp Ser Phe Phe Val Ala Leu Asp Met Xaa Asp Gly Thr
          35           40           45
Leu Ser Phe Ile Val Asp Gly Gln Tyr Met Gly Val Ala Phe Arg Gly
          50           55           60
Leu Lys Gly Lys Lys Leu Tyr Pro Val Val Ser Ala Val Trp Gly His
65           70           75           80
Cys Glu Ile Arg Met Arg Tyr Leu Asn Gly Leu Asp Pro Glu Pro Leu
          85           90           95
Pro Leu Met Asp Leu Cys Arg Arg Ser Val Arg Leu Ala Leu Gly Lys
          100          105          110
Glu Arg Leu Gly Ala Ile Pro Ala Leu Pro Leu Pro Ala Ser Leu Lys
          115          120          125
Ala Tyr Leu Leu Tyr Gln
          130

```

(2) INFORMATION FOR SEQ ID NO:42:

- 165 -

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 265 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

```

AAGGGTAAAA AACTGTATCC TGTAGTGAGT GCCGTCTGGG GCCACTGTAG ATCCGAATGC      60
GCTACTTGAA CGGACTCGAT CCCGAGACTG CCGCTCATGG ATTTGTGCCG TCGCTCGGTG      120
CGCCTGGCCC TGGGGAGGGA GCGCCTGGGG GAGAACCACA CCTGCCGCTG CCGGCTTCCC      180
TCAAGGCCTA CCTCCTCTAC CAGTGACGTT CGCCATCATA CCGCCAGCGC GACAGCCACC      240
TGGTGCCAAC TCACTGAGCC GCCTG                                     265

```

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2438 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

```

AAGTGGCGGC GGTCCCTGGA GAGCAGGCGG AGGCAGCGGC AAGTCTGACT CTGGGCTGAC      60
CGTGGAGCCG GGGCGGGGGC TGACAGCCAG GCCTCCGCCT GGCGGGAGCC GCACGAGGAG      120
CGGGAGTGGC CGGGCCTCTC TTCCGCGCTT GAGCGAGCGC CGGGTGATGG CGGTGGTGAT      180
GGCGGCAGGC GCTCGGACAG CTCCGCTTGA GCTGAGCTCG GAGAGATCCG TCCAGAAAGT      240

```

- 166 -

GCCCAGAAGA AACTTCCTCT TAGAAAAGCT GAAAAACACA RTATTTATAA CACTGGAAAT	300
TGTAAAGAAT TTGTTTAAAA TGGCTGAAAA CAATAGTAAA AATGTAGATG TACGGCCTAA	360
AACAAGTCGG AGTCGAAGTG CTGACAGGAA GGATGGTTAT GTGTGGAGTG GAAAGAAGTT	420
GTCTTGGTCC AAAAAGAGTG AGAGTTGTTC TGAATCTGAA GCCATAGGTA CTGTTGAGAA	480
TGTTGAAATT CCTCTAAGAA GCCAAGAAAG GCAGCTTAGC TGTTTCGTCCA TTGAGTTGGA	540
CTTAGATCAT TCCTGTGGGC ATAGATTTTT AGGCCGATCC CTAAACAGA AACTGCAAGA	600
TGCGGTGGGG CAGTGTTTTT CAATAAAGAA TTGTAGTGGC CGACACTCTC CAGGGCTTCC	660
ATCTAAAAGA AAGATTCATA TCAGTGAACAT CATGTTAGAT AAGTGCCCTT TCCCACCTCG	720
CTCAGATTTA GCCTTTAGGT GGCATTTTAT TAAACGACAC ACTGTTCCCTA TGAGTCCCAA	780
CTCAGATGAA TGGGTGAGTG CAGACCTGTC TGAGAGGAAA CTGAGAGATG CTCAGCTGAA	840
ACGAAGAAAC ACAGAAGATG ACATACCCTG TTTCTCACAT ACCAATGGCC AGCCTTGTGT	900
CATAACTGCC AACAGTGCTT CGTGACAGG TGGTCACATA ACTGGTTCTA TGATGAACTT	960
GGTCACAAAC AACAGCATAG AAGACAGTGA CATGGATTCA GAGGATGAAA TTATAACGCT	1020
GTGCACAAGC TCCAGAAAAA GGAATAAGCC CAGGTGGGAA ATGGAAGAGG AGATCCTGCA	1080
GTGGAGGCA CCTCCTAAGT TCCACACCCA GATCGACTAC GTCCACTGCC TTGTTCCAGA	1140
CCTCCTTCAG ATCAGTAACA ATCCGTGCTA CTGGGGTGTC ATGGACAAAT ATGCAGCCGA	1200
AGCTCTGCTG GAAGGAAAGC CAGAGGGCAC CTTTTTACTT CGAGATTCAG CGCAGGAAGA	1260
TTATTTATTC TCTGTTAGTT TTAGACGCTA CAGTCGTTCT CTTGATGCTA GAATTGAGCA	1320
GTGGAATCAT AACTTTAGCT TTGATGCCCA TGATCCTTGT GTCTTCCATT CTCCTGATAT	1380
TACTGGGCTC CTGGAACACT ATAAGGACCC CAGTGCCTGT ATGTTCTTTG AGCCGCTCTT	1440
GTCCACTCCC TTAATCCGGA CGTTCCCCTT TTCCTTGAG CATATTTGCA GAACGGTTAT	1500
TTGTAATTGT ACGACTTACG ATGGCATCGA TGCCCTTCCC ATTCCTTCGC CTATGAAATT	1560
GTATCTGAAG GAATACCATT ATAAATCAAA AGTTAGGTTA CTCAGGATTG ATGTGCCAGA	1620

- 167 -

GCAGCAGTGA TCGGAGAGG TTAGAATGTC GACCTGCATA CATATTTTCA TTTAATATTT	1680
TATTTTTCTT ATGCCTCTTT GAATTTTTGT ACAAAGGCAG TTGAATCAAA TAAAACTGTG	1740
CCCTAAGTTT TAATCCAGA TCAATTTATT TTTTTATGA TACACTTGTT ATATATTTTT	1800
AAGCAGGTGT TTGGTTTTGT TTTTACCATA TAAATTTACA TATGGTCCAG GCATATTTAC	1860
AATTTCAAGG CATTGCATAT ACATTTGAAT ATTCTGTATT TTTTAAATAA TCTTTTGTTT	1920
TTTCCTATGT GTGAAATATT TTGCTAATCT ATGCTATCAG TATTCCTGTA TGACCGAATA	1980
GTTACCTATT CTCTTTTCAT CTTGAAGATT TTCAGTAAAG AGTGTTGTAA TCAATCCATT	2040
ATAATGTAAT TGACTTTTGT AATTTGCCAA TAGGAGTGTT AAACAACAAA ATGATTTAAA	2100
ATGAACTTA ATGTATTTTC ATTTTAAATA TTAATAAAC CAAGTTTGTT TGTAGTTAT	2160
TCTAGCCAAT AAGAAAAGAG AATGTAGCAT CCTAGAGGTG TATTTGTTCT GCAGTTTGGC	2220
AGGACCGTCA GTTAGTCCAA ATAAACATCC CCTCAGCGTG GAGGCGAATG GAACCTGTGC	2280
TCCTTTCTTA CGGGAAGCTT TGCAAAGCAA AATAGCAGGG TTACAAGCTT GGAGTTGTTA	2340
AGGCAACTAG AGTTTTCTCT ATTAATTTAT AGACTGTTGT TGCACCTACT TAGCTCTTTT	2400
TTGGGAACTC TAGTTCCCAG GGGAAAATAC CTCGTGCC	2438

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 542 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Ser	Gly	Gly	Gly	Pro	Trp	Arg	Ala	Gly	Gly	Gly	Ser	Gly	Lys	Ser	Asp
1				5				10				15			

- 168 -

Ser Gly Leu Thr Val Glu Pro Gly Arg Gly Leu Thr Ala Arg Pro Pro
 20 25 30

Pro Gly Gly Ser Arg Thr Arg Ser Gly Ser Gly Arg Ala Ser Leu Pro
 35 40 45

Arg Leu Ser Glu Arg Arg Val Met Ala Val Val Met Ala Ala Gly Ala
 50 55 60

Arg Thr Ala Pro Leu Glu Leu Ser Ser Glu Arg Ser Val Gln Lys Val
 65 70 75 80

Pro Arg Arg Asn Phe Leu Leu Glu Lys Leu Lys Asn Thr Xaa Phe Ile
 85 90 95

Thr Leu Glu Ile Val Lys Asn Leu Phe Lys Met Ala Glu Asn Asn Ser
 100 105 110

Lys Asn Val Asp Val Arg Pro Lys Thr Ser Arg Ser Arg Ser Ala Asp
 115 120 125

Arg Lys Asp Gly Tyr Val Trp Ser Gly Lys Lys Leu Ser Trp Ser Lys
 130 135 140

Lys Ser Glu Ser Cys Ser Glu Ser Glu Ala Ile Gly Thr Val Glu Asn
 145 150 155 160

Val Glu Ile Pro Leu Arg Ser Gln Glu Arg Gln Leu Ser Cys Ser Ser
 165 170 175

Ile Glu Leu Asp Leu Asp His Ser Cys Gly His Arg Phe Leu Gly Arg
 180 185 190

Ser Leu Lys Gln Lys Leu Gln Asp Ala Val Gly Gln Cys Phe Pro Ile
 195 200 205

Lys Asn Cys Ser Gly Arg His Ser Pro Gly Leu Pro Ser Lys Arg Lys
 210 215 220

Ile His Ile Ser Glu Leu Met Leu Asp Lys Cys Pro Phe Pro Pro Arg
 225 230 235 240

Ser Asp Leu Ala Phe Arg Trp His Phe Ile Lys Arg His Thr Val Pro
 245 250 255

- 169 -

Met	Ser	Pro	Asn	Ser	Asp	Glu	Trp	Val	Ser	Ala	Asp	Leu	Ser	Glu	Arg			
			260					265					270					
Lys	Leu	Arg	Asp	Ala	Gln	Leu	Lys	Arg	Arg	Asn	Thr	Glu	Asp	Asp	Ile			
	275						280					285						
Pro	Cys	Phe	Ser	His	Thr	Asn	Gly	Gln	Pro	Cys	Val	Ile	Thr	Ala	Asn			
	290					295					300							
Ser	Ala	Ser	Cys	Thr	Gly	Gly	His	Ile	Thr	Gly	Ser	Met	Met	Asn	Leu			
305					310					315					320			
Val	Thr	Asn	Asn	Ser	Ile	Glu	Asp	Ser	Asp	Met	Asp	Ser	Glu	Asp	Glu			
				325					330				335					
Ile	Ile	Thr	Leu	Cys	Thr	Ser	Ser	Arg	Lys	Arg	Asn	Lys	Pro	Arg	Trp			
		340						345					350					
Glu	Met	Glu	Glu	Glu	Ile	Leu	Gln	Leu	Glu	Ala	Pro	Pro	Lys	Phe	His			
	355					360						365						
Thr	Gln	Ile	Asp	Tyr	Val	His	Cys	Leu	Val	Pro	Asp	Leu	Leu	Gln	Ile			
	370					375					380							
Ser	Asn	Asn	Pro	Cys	Tyr	Trp	Gly	Val	Met	Asp	Lys	Tyr	Ala	Ala	Glu			
385				390						395					400			
Ala	Leu	Leu	Glu	Gly	Lys	Pro	Glu	Gly	Thr	Phe	Leu	Leu	Arg	Asp	Ser			
			405					410					415					
Ala	Gln	Glu	Asp	Tyr	Leu	Phe	Ser	Val	Ser	Phe	Arg	Arg	Tyr	Ser	Arg			
		420						425					430					
Ser	Leu	His	Ala	Arg	Ile	Glu	Gln	Trp	Asn	His	Asn	Phe	Ser	Phe	Asp			
	435						440					445						
Ala	His	Asp	Pro	Cys	Val	Phe	His	Ser	Pro	Asp	Ile	Thr	Gly	Leu	Leu			
	450					455						460						
Glu	His	Tyr	Lys	Asp	Pro	Ser	Ala	Cys	Met	Phe	Phe	Glu	Pro	Leu	Leu			
465				470					475					480				
Ser	Thr	Pro	Leu	Ile	Arg	Thr	Phe	Pro	Phe	Ser	Leu	Gln	His	Ile	Cys			
			485					490						495				

- 170 -

Arg Thr Val Ile Cys Asn Cys Thr Thr Tyr Asp Gly Ile Asp Ala Leu
 500 505 510

Pro Ile Pro Ser Pro Met Lys Leu Tyr Leu Lys Glu Tyr His Tyr Lys
 515 520 525

Ser Lys Val Arg Leu Leu Arg Ile Asp Val Pro Glu Gln Gln
 530 535 540

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4999 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CCCTCTGGGC AAGCCGCCCC CCCCCACCC ATCTACCACA CACACACACA CACACACACA	60
CACACATTCA GACCTTGGGG CAAAAACAAA GCAAATAAC AACAACAAAA AACTGCCTG	120
TGGAAAGTCC TTA CTTCAGG AAGGTTGGCA GATGAGGAGC AAGGGAACAT TTTATCAGGA	180
CTGCCACAAA GGAGCTTTTT TTTTAAATGG TTTTCAAGA CAGGGTTTCT CTGTATAGCC	240
CTGGCTGTCC TGGAGCTCAC TTTGTAGACC AGGCTGGCCT CGAACTCAGA AATTCGCCTG	300
CCTCTGCCTC CTGAGTGCTG GGATTAAAGG CGTGCAGCAC CATGTCCAAC TGGCATT TTC	360
TCAATTAAGG TTCGTTCTT TCAGATAACT CTAGGTTCTG GGTCAAGCTG ACACAAGGCT	420
ACACAGCACA GTTTGTATGC CACATTCAGT TCAGAAGACA CCCAACCTCC CTGGAAGTGG	480
AACTTATGCA CATTTGTGAG CTTCCTCTTG GGAGTGGGAA CCTGAACTGG GTCCTCTGCA	540
AGAGCAGCCG TGCTCTTAAC TGCTGAGCCA TTCAGCAGC CTCACATCAG AATTAAGTTA	600

- 171 -

GAAATTAGCCG GGTATGAATC ATACCCTTAG AATCCTAGCA TCTGAAAGCA GAGCTAAGAG
660

AAACAGGGAT TCAAGACCAG CTCTTGGCTA CAGAGCCCGT CCTGTCCTAG GATGGGCTAC 720

AAGAGACTAT TTCAAAGCCA TCCAAACAAC AATAACTACA ACAACAACAA GGTAAAAATT 780

AGGCTGGGCA CAGGGTACAC ACCTTTAATG CCAACACTCA GGAGGCAGAG GCAGGCTGAT 840

CAGTGTGAGT TTGAGTTCAA CGTGGTCTAC ATAGGGAGTT CTAGGCCAGC AGAGGTTACA 900

GTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCACACA CACACACACA CACACACACA 960

CACACACACA CACACACGGT GGCATTATGG GATTTTTTTG GGATAAGGTT TCTCTGTCTA 1020

GCCCTGGCAT AGATTCACCTC TGTAGACTAG GCTAGCCTTG AACTCAGAGA TCCGCCTGCC 1080

TCTGCCTCCC AAGTGCTGGG ATTATAGGTG TTGCACCACC ACTGCCCAGC CACTTTGGGA 1140

TTTTTGAAGT GTTATCAAGA GGCTTTCGAG GAGGTCAAAC TTCAACAGCA ACCTCTCCAT 1200

GATAATGTAG CTAATGATCA AACGACACTC AAAACTTAAC CCTTAAAGCA CACATCCACC 1260

AGACAGCGTG CCCACTCGTA GTTCCATTAC TCAGGAGGCT GAAGCAGGAG GATGAAGGAC 1320

TAAGGCTTCA GCAACCTAGG GAGCCGCAGG GGACAGTAGT CTCAATCCCT ACATTCTCCT 1380

GAACACAGGA GCAGGAGTTC AGGAAGGGTG TCAAGGCCGC TTAAGTATCT TAGGGCCTCA 1440

GGAATGACTA GCTCAGGCAG AGAGAGCAAA GGTCTCCAGT GGAGAAGTCT ACACACACAC 1500

ACACACACAC ACACACACAC ACACACACAC AGAATCCAAG GCGATGACGT CATCAAAGGG 1560

TTAATTCTAG TCTGGGATGG GGGGGAGGGT GGGGCACGCA GCTGTCAGGT GGCTTTGGAA 1620

AAATAAACTG CTGAAGAGTC TGACGCCAGG GAGTCCTGGG AGGGACAAGA GGTTACCCAC 1680

TCAAAGAGTG TGCTCCACAA AGCATGCGCG CTTGTCCACG TCTGGAGTCG TCACTTATTT 1740

TTTGCCTGGA TTCTTTGTAG CCGGTGGGTT CTCAAGGCGG TAAGTGGTGT GGCCGCCGTG 1800

GTCTGGGAGG TGACGATAGG GTTAATCGTC CACAGAGCCC AGGGGCGGAG CGCGGGCGGG 1860

CGTCCGCAGC CCCGCTGGAG CCGGAAGCAG TGGCTGGTCA GGGGCGCTTC TAGCCTTCCC 1920

- 172 -

TATCTGTACT	TCCACAGAGG	TCTCTGCGAG	CTAGGGGGAC	AGTGAGGTGC	GGGGTAGGGG	1980
CCCGGCGTTA	GAGCCAGCAA	GGGGACGGTT	CACGGTAAGG	TCTGAGGGAG	AGAGAGCTCC	2040
TGAGAAACTT	GGGGGGCGCG	ACACAGATAG	GGTGAAAGCA	GAGTGATAGA	CCTGGGATGG	2100
TTAGGGGACC	AAGGGAAGAC	CAGGCTGGTT	GGCATACACC	GGTGAACGGA	TGGGAGTCCT	2160
AGGGAAGAT	GATGCGCCTA	ACAGTCCTTT	CTGTCTCCAC	ACCACTCCAG	GGGACGATCC	2220
GGAGCTCAAC	TTTCAAAAGC	GAGACGCCCC	AGCAAGCCTG	TTTTGAGAAG	TTCTTCAGCG	2280
GCTCTCCTCA	TGGGCCAGAC	GGCCCTGGCA	AGGGGCAGCA	GCAGCACCCC	TACCTCGCAG	2340
GCTCTGTACT	CGGACTTCTC	TCCTCCCGAG	GGCTTGGAGG	AGCTCCTGTC	TGCTCCCCCT	2400
CCTGACCTGG	TTGCCCAACG	GCACCACGGC	TGGAACCCCA	AGGATTGCTC	CGAGAACATC	2460
GATGTCAAGG	AAGGGGTCT	GTGCTTTGAG	CGGCGCCCTG	TGGCCCAGAG	CACTGATGGA	2520
GTCCGGGGGA	AACGGGGCTA	TTCGAGAGGT	CTGCACGCCT	GGGAGATCAG	CTGGCCCCTG	2580
GAGCAAAGGG	GCACACACGC	CGTGGTGGGC	GTGGCCACCG	CCCTCGCCCC	GCTGCAGGCT	2640
GACCACTATG	CGGCGCTTTT	GGGCAGCAAC	AGCGAGTCCT	GGGGCTGGGA	TATTGGGCGG	2700
GGAAAATTGT	ATCATCAGAG	TAAGGGCCTC	GAGGCCCCCC	AGTATCCAGC	TGGACCTCAG	2760
GGTGAGCAGC	TAGTGGTGCC	AGAGAGACTG	CTGGTGGTTC	TGGACATGGA	GGAGGGGACT	2820
CTTGGCTACT	CTATTGGGGG	CACGTACCTG	GGACCAGCCT	TCCGTGGACT	GAAGGGGAGG	2880
ACCCTCTATC	CCTCTGTAAG	TGCTGTTTGG	GGCCAGTGCC	AGGTCCGCAT	CCGCTACATG	2940
GGCGAAAGAA	GAGGTGAGAT	ACGGACTAGG	TGTGGGGAGA	TCACTACTCT	TGGCAATGGT	3000
TTGGGCTGGA	AACTCATGGT	TGGAGCACAG	GAAGTAGGCT	TCTTGTCACT	TTGGCCTGTC	3060
ACTTAGATGG	CCTTGGATCT	AGCTTCACTC	CCAATCCCTA	TTGGATGTGA	TGCACAAATT	3120
CAGAGCCTTT	GGGTCTCCCT	CAGCTGAGGT	GGCGGTGGAA	ATGGAGGAAG	AAGGAAGGGT	3180
GCCTGAGCAG	GATCTCAAGT	TCAAGGATGC	CTGGAGTTGC	TTACTTACCT	TGTCTTCCTT	3240
CTCTCTCCGC	AGTGGAGGAA	CCACAATCCC	TTCTGCACCT	GAGCCGCCTG	TGTGTGCGCC	3300

- 173 -

ATGCTCTGGG	GGACACCCGG	CTGGGTCAAA	TATCCACTCT	GCCTTTGCCC	CCTGCCATGA	3360
AGCGCTATCT	GCTCTACAAA	TGACCCAGTA	GTACAGGGTG	TGCTGGCACC	CTACCGTGGG	3420
GACAGGTGGA	GAGGCACCCG	CTGGCCTAGA	CAACTTTAAA	AAGCTGGTGA	AGCTGGGGGG	3480
GGGGGGCTGG	ACCCCTTCAC	CTCCCCTTCT	CACAGGAGCA	AGACATATAG	AAATGATATT	3540
AAACACCATG	GCAGCCTGGG	ACAAAGAGGT	TTTTGAAGTA	AAAAATGAGA	TGTATTGTCA	3600
CAACCTGTTT	CATTATTGTT	TTTTGTTTTG	TTTTACACTC	CCCCACCCCA	GGCTAGAGCC	3660
CCATCACTGT	CTTAAGGAAT	TATGACAACC	CACAAAGCTC	AGGCCCAGGT	GTTTATTTCC	3720
CTTACATGTA	GGATGGTTCA	CAAACACAAT	ACAGGGGCTT	TGGCACCGTG	GGGGAGGGGA	3780
CTATCCCAGG	CCTCTTAGGG	TCTCATGTAT	ACCGAATTCA	GACCCGAAAG	CTCTGAATTT	3840
CTGCATCAGA	CATCCAGTAG	AACTTGGGAG	TGAAGCTAGA	GCCAAGGCCA	TCTAAGTGAC	3900
AGGCCAAAGT	GACACGAAGC	CCACTTCCTG	TGCTCCAACC	ATGAGTTTCC	AGCCCAAACC	3960
AATGGAAGGT	GATTTCACTT	GTCAGGGCCC	AAAGGGACAG	TCAGTTCTAC	TCCCTCCCCT	4020
CACTAGGAGC	CACCTTGGTG	ACAGTTGATT	CTACCCACTG	TAAGTGGTAA	AGGGATTGGC	4080
CTGGTCCCAA	CCATAATAGG	GCGGTGGAAA	CGGCTCAGGA	GGGTACAGCG	TGGATTAGGC	4140
CACAAGATGG	GGCAGATGAT	GTCATCAGAA	GCATGTGACC	GGTGGGAGCA	GTTACTAAAC	4200
TTCTGGGCAA	CCTAGTCCAT	GCTATGCAGG	CAGGTAGAGG	GATGGGCAGT	GCTCATTGTT	4260
TGGCATTGAT	GATGTCCACA	AATTCAGGCT	TGAGAGATGC	GCCACCCACA	AGGAAGCCGT	4320
CCACGTCAGG	CTGGCTTGCC	AGCTCTTTGC	AGGTTGCTCC	AGTCACAGAA	CCTGTACCAG	4380
GAACAAGAAG	ACAGTTTGGT	CAGGTCTATG	ATCAGAACAC	TTAAGCCCCA	CCTCTCTGTG	4440
CAAGGCAGCC	TCAGTCTGTC	TTAGCCCATT	TCCGTCTTAG	CTAGAGCCAA	AGCCACTCAC	4500
CTCCATAAAT	GATCCGGGTG	CTCTGAGCCA	CCCCATCATT	GACATTGGAT	TTCAGCCATC	4560
CCCGGAGCTT	CTCGTGTA	CTCCTGTGCCT	AGAAGGAGGA	GGCAGAGCTA	CTAAGTAAGC	4620
TCCTTCCTAT	CTATCATTCA	AGGAGTAAAA	ACCACTGGTT	CTCACATAGA	GTTGAGTTTC	4680

- 174 -

CAGAAAAGCC CCGGGACCAG AGAGTGGCAA GGCTCCAATC CCACCAGGCT TGAATGAAC 4740
 ATTTTGGCA AAGTCACTCT CCTGGTGAG TTTGGGGGCC CTCTGTCTCT AAAGGGGCTT 4800
 GGATGGGCTC CATAGCTGTG TGAGTCTGTT AAAGCCGGAC AGGCTGAGGA GCTCTGGGTA 4860
 GTTACCTGCT GAGGGGTTGC CGTCTTGCCA GTCCCAATGG CCCACACAGG TTCATAGGCC 4920
 AGGACCACCT TGCTCCAGTC TTTCACATTA TCTGTGGGGC AGAGAGGAGA GTGAGTAGGA 4980
 AGGAGCTGAC CCGCCAAGC 4999

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 264 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Met Gly Gln Thr Ala Leu Ala Arg Gly Ser Ser Ser Thr Pro Thr Ser
 1 5 10 15
 Gln Ala Leu Tyr Ser Asp Phe Ser Pro Pro Glu Gly Leu Glu Glu Leu
 20 25 30
 Leu Ser Ala Pro Pro Pro Asp Leu Val Ala Gln Arg His His Gly Trp
 35 40 45
 Asn Pro Lys Asp Cys Ser Glu Asn Ile Asp Val Lys Glu Gly Gly Leu
 50 55 60
 Cys Phe Glu Arg Arg Pro Val Ala Gln Ser Thr Asp Gly Val Arg Gly
 65 70 75 80
 Lys Arg Gly Tyr Ser Arg Gly Leu His Ala Trp Glu Ile Ser Trp Pro
 85 90 95
 Leu Glu Gln Arg Gly Thr His Ala Val Val Gly Val Ala Thr Ala Leu

- 175 -

100	105	110
Ala Pro Leu Gln Ala Asp His Tyr Ala Ala Leu Leu Gly Ser Asn Ser		
115	120	125
Glu Ser Trp Gly Trp Asp Ile Gly Arg Gly Lys Leu Tyr His Gln Ser		
130	135	140
Lys Gly Leu Glu Ala Pro Gln Tyr Pro Ala Gly Pro Gln Gly Glu Gln		
145	150	155
Leu Val Val Pro Glu Arg Leu Leu Val Val Leu Asp Met Glu Glu Gly		
165	170	175
Thr Leu Gly Tyr Ser Ile Gly Gly Thr Tyr Leu Gly Pro Ala Phe Arg		
180	185	190
Gly Leu Lys Gly Arg Thr Leu Tyr Pro Ser Val Ser Ala Val Trp Gly		
195	200	205
Gln Cys Gln Val Arg Ile Arg Tyr Met Gly Glu Arg Arg Val Glu Glu		
210	215	220
Pro Gln Ser Leu Leu His Leu Ser Arg Leu Cys Val Arg His Ala Leu		
225	230	235
Gly Asp Thr Arg Leu Gly Gln Ile Ser Thr Leu Pro Leu Pro Pro Ala		
245	250	255
Met Lys Arg Tyr Leu Leu Tyr Lys		
260		

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5615 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- 176 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GTACTTTCTT TATATCTCCA TAATTTTATT TACTATTACT ACATGATACA TTATTTTATA	60
AAAGTCTTTG TAACCTCCTT AAGGATTCAC TGCTTAATCT CCAGTGCTTA GCACAAATCA	120
TTAAATGCGA ACCAGAACT CTTCCAAATG TGTACATCT ATAACCTCAT TGGATTCTCA	180
CTACCAACCC CATGCAATAG ATACTAATGT GATCTCTGTC TTACAGAGGA AGAAACAGGC	240
ACAGGGAGGT TCAGTAATTT GCCCAAGGTC ATACACACAC TGGCCTTCAG GTATTCATGC	300
CCGGGGAGTC TGGTCCCACA GCTGGCATGT TTGCCATTAT ATTATATTGC CTCCTTATAG	360
TGTCGGCACT CATTAAGCAC ATTGACAGCT ATGCTTGGTG AGTGACTACT ATGTACCCAG	420
CTCTGTGCTA CATGCTTTAC CTGGATTATT TCAACTGCAC AACAACCCTG TGAGGTAAC	480
ACCATCATTG CTCCTATTTT ACATAACAGA AACTTACAGA AATCTGGGGC TGGGCGTAGT	540
GGCTCATGCC TGAAATCCCA GCACTTTGGG AGACCCTGTC TCTAAAAAA ATTTTTTTTT	600
GGCCGGACGT GGTGGCTCAC ACCTGTAATC TCAGCACTTT GGGAGGCTAA GGCAGGCAGA	660
TCACAAGGTC AGGAGTTCTA GACCAGCCTG GCCAACATGG CAAAACCCTG TGTCTACTAA	720
AAATACAAAA AATAGCTAGG CGTGGTGCCA GGTGCCTGTA ATCCCAGCTA CTCAGGAGGC	780
TGAGGCAGGA GAATCCCCTG AACCTGGGAG ATGGAGGTTA CAGAGAGCCG AGATCGTGCC	840
GCTGCACTCC AGCCTGGGCA ACAAGAGCAA GACTCTGTCT CGAAAAAAT AAAAAATAAA	900
ATAAAAAATAT TTTTTTAAAA ATTAGCTGGG TGTGGTAGCA CATGCCTGTA GTCCCAGCTA	960
CTTGGGAGGC TGAGGTAGGA GGATCACTTG AGCCCAGGAG GTCAAGGCTG CAGTGGGCTG	1020
TGATGGCGCC ACTGCACTCT AGCCTTGGTG ACAGCAAGAC CCTGTCTCAA AAAAAAAAAA	1080
AAGAGAAATC GGGCAACTTC CCCAAGATCG CGCAGTTAAC TAGTGGCATA GCTTCACTCA	1140
AACTCGAAGT CTTAATCAGG AACTCTACC AAATGAGATC AACGGCTCAG TAATGGATTG	1200
GCATCCAGTA TGAAGACTGG ACCAGCAGGG AGAACTATGA TCGGTACAGC CTAGAGCCTG	1260
AAGCAGATTT CACAGCCTCA GAGGTGGCAC AGGCTGACTC ACAACCCGGG GCAGAAAGGG	1320

- 177 -

ACCAGCCCAG AACAGTGAC CCAGAATCAC AGGGAAGTAG AAATGGGATT CGGCACAATG	1380
AAGCCCCTCC TTGACCCCAT GCTCCTTACC CTCAGGGGCG CAGGAGTTAG TCGCTCAGGC	1440
GGCTCAAAGG TCTTGACGGT GGAGAACACC ATCCCCAGGG ATTCCCGACG CGGTGATGCC	1500
ATCAAAGCGT TAATTCTGAG ATGGGCCTGC CCGGGTGCGG ACTCTGCCGC AGCAAGAGAA	1560
GGGTAACTG CCCCGGGCCT TCGCCGTGGG GCGGGGGCCT CGGGGAGGGT CACAGCCCCG	1620
GACTGAGACC CGAGGTAAAC CGCCCGGGGT GGGCTCCACG GGGGCGGGG ATGCTCTCCG	1680
CGGCTGCTGC CGGTATAGAG CGGTAAGTGC CCAGGAGGGG GCGGGGCCCC ACAGGGGCGT	1740
GGCCTCGGAG CTGCACGGCC GTGGGCGGCG ATGAGAGGGT TAAGCCCCAG AGGGCCCTGG	1800
AGGGGCGGGG CCGCGGGACG GGCTCGGCCC AAGGGAGGAG CTGGGGGCGG AAGCGCCCGG	1860
CGGTCTGCGC CCTGCGCGCC TCGGCTTCTT TCCGCCCGGC TCCTTCAGAG GCCCGGCGAC	1920
CTCCAGGGCT GGAAGTCAA CCAGGTTTCG GGGGCAGCGG CGAGGGCTCC GGGCGAGTAA	1980
GGGGGATGGT CCATGCTGAG GCCCAAATGG GGCGAACTCG CGAGAGTCTC TGGCGACCTG	2040
GATCAGATGG GGCGAGGGCA GATGAAGGGC CCAGGAGCTT TGGGGCAGCG AGGAGGGAGG	2100
AGCGGGCCCC TTGGCAAAC TGGGTGAAAG GATGGGGTAC CTGGGTGACG AGCCCCCGCC	2160
AGGATTTCTGC TCTTCACGCC CCTTTTCTCC CAGCTCCCTT CCAGGTCAAT CCAAAC TGGA	2220
GCTCAACTTT CAGAAGAGAA AGACGCCCCA GCAAGCCTCT TTCGGGGAGT CCTCTAGCTC	2280
CTCACCTCCA TGGGCCAGAC AGCTCTGGCA GGGGGCAGCA GCAGCACCCC CACGCCACAG	2340
GCCCTGTACC CTGACCTCTC CTGTCCCGAG GGCTTGGAAG AGCTGCTGTC TGCACCCCT	2400
CCTGACCTGG GGGCCAGCG GCGCCACGGT TGGAACCCCA AAGACTGTTC AGAGAACATC	2460
GAGGTCAAGG AAGGAGGGTT GTACTTTGAG CGGCGGCCCC TGGCCAGAG CACTGATGGG	2520
GCCCGGGGTA AGAGGGGCTA TTCAAGGGGC CTGCACGCCT GGGAGATCAG CTGGCCCTA	2580
GAGCAGAGGG GCACGCATGC CGTGGTGGGC GTGGCCACGG CCCTCGCCCC GCTGCAGACT	2640
GACCACTACG CGGCGCTGCT GGGCAGCAAC AGCGAGTCGT GGGGCTGGGA CATCGGGCGG	2700

- 178 -

GGGAAGCTGT ACCATCAGAG CAAGGGGCCC GGAGCCCCC AGTATCCAGC GGGAAGCTCAG	2760
GGTGAGCAGC TGGAGGTGCC AGAGAGACTG CTGGTGGTTC TGGACATGGA GGAGGGAAGT	2820
CTGGGCTACG CTATTGGGGG CACCTACCTG GGGCCAGCAT TCCGCGGACT GAAGGGCAGG	2880
ACCCTCTATC CGGCAGTAAG CGCTGTCTGG GGCCAGTGCC AGGTCCGCAT CCGCTACCTG	2940
GGCGAAAGGA GAGGTGAGGC CTGGGGCAGA CGTGGGGAGA ACTTTCTGTC CCTGGTGGCA	3000
GTGGTTTGGG ATGGAAGTCT TTCTGACAAG AGCAGAGGGG ATGGACCTTC ATCCAGCCTG	3060
CCTCAACCTC TGTTCACTGC TGGGAAAGGC TAGGGGTCTT CACAGCTGTT ATTTAATTTA	3120
ACCCAACAGC AATAGAGGTG AAACAGGCTT GAGAAAGCAA CTTTCTCAAG TTCTCTGGC	3180
CAGTAAATGG TGAACCTTCA GAATGGAGGG AGGAACTGCA GGGATGAGAG AATTCAGGAG	3240
ATATCAACCC CTGAGCAAGA GGTGCAAAGC GTTAGGTACT GGGTTTGATG TACAGGTCCA	3300
AAAGAAGGAT GGGCAGAGCC AGGTACCCAG GCTGTATACC GGATCCCTG GGCTCTAACC	3360
TGTCTCTGTG CCACATACCT ACTTCCTTCC TCAGCCACAC CTCTGGATGG AGACACTGGG	3420
GCCCTGGGCA CCAGGGAGGA GAGCAGTGGA GGAGGCAGGG CCTTAGGGTG GGGCAGCAGG	3480
GGAGGAGCCT CCCAGGAAC TGAAGGGTC CAGGGCTTGG AGCTGCTCTC TGCAGTTGTG	3540
TGGGCTGTAG AGTGGAGGGC CATCCCTCCT CACCTCAGCC CCAGCTCCCA AGCCTCTGGA	3600
GTCAAAGCCT GGGCCAGCTC CACCACTGTC AGAGCCACCT TGGCCTGTTG TTTAGAGGGC	3660
CTTAGCCAGC TCTTCACCCC CAGCTCTGAC TAGGGATGTG TGAAATCTTA TCTGGGAGGC	3720
AGAACTTCCG GGTATCTCAA ATTCCCTTT CAGCCAGGTG GGCACACTCG AAGCAGGAAA	3780
GCAGAAAGGC ATCTGAGTAG GACCCCGTAG TTTGAGGACA TCTGGCTGGT GGCTGCACCC	3840
ATACTTACAT TCCCTCCTT CTCTCTCCCA GCGGAGCCAC ACTCCCTTCT GCACCTGAGC	3900
CGCCTGTGTG TGCGCCACAA CCTGGGGGAT ACCCGGCTCG GCCAGGTGTC TGCCCTGCCC	3960
TTGCCCCCTG CCATGAAGCG CTACCTGCTC TACCAGTGAG CCCTGTGATA CCACAGACTG	4020
TGCTGAGGTC TTGCCACCAC CCCTCCCCTT GGGGAGGTGG GGAGGCACTG CTGGCCTAGA	4080

- 179 -

CCAGCTGCTG AAAGCTGGTG AGGCTGAGCC CCTACCCCAA CCCAAGCTCT GCGGAAATCA	4140
ACAGCCCCAG AGCCACTTGG AGGGAGGAAG AAAGGGAGCC GGCCTTCAAG GCTATGACAG	4200
TCTGCTACGC AAAACATTTT TTCAAGTAAA AATAGTAAGA GATGTTGTTA TAGAAACCTG	4260
TTCTTGTTTT TTTTTTTTTC TTGCACAAAT GATCATTTAT ATAGCTGCCT CAAAAAGGAA	4320
GATTATCTGG GCAAGTCCAG TGAAGGCAGA CAAACCACAA GACCTAGTGC CAGGTTTATT	4380
CCCTCACATG GGTGGTTCAC ATACACAGCA CAGAGGCACG GGCACCATGG GAGAGGGCAG	4440
CACTCCTGCC TTCTGAGGGG ATCTTGCCCT CACGGTGTA GAAGGGAGAG GATGGTTTCT	4500
CTTCTGCCCT CACTAGGGCC TAGGGAACCC AGGAGCAAAT CCCACCACGC CTTCCATCTC	4560
TCAGCCAAGG AGAAGCCACC TTGGTGACGT TTAGTTCCAA CCATTATAGT AAGTGGAGAA	4620
GGGATTGGCC TGGTCCCAAC CATTACAGGG TGAAGATATA AACAGTAAAG GAAGATACAG	4680
TTTGATGAG GCCACAGGAA GGAGCAGATG ACACCATCAG AAGCATATGC AGGGAAAGGG	4740
CAGTTACTGG GCTTCTGGGC TGCTTAGTCC CTGGCTTGGC AGGAAGGGTA GGGAAGATGG	4800
ATGGGGCTCA TTGTTTGGCA TTGATGATGT CCACGAATTC GGGCTTGAGG GAAGCACCAC	4860
CCACAAGGAA GCCATCCACA TCAGGCTGGC TGGCCAGCTC CTTGCAGGTT GCCCCAGTCA	4920
CAGAGCCTGG GAAGGGAGCA GAACAAGGGC TTGGTCAAGA ATGGGATGAG TCTGCCCCAT	4980
CCCCACCTCC ATGTCCGAGG GCTCAGTCTA GTCCTCAGCC CACTCCACCT CAGCCGGGAA	5040
CCAAAGCCAC TCACCTCCAT AAATGATACG GGTGCTCTGA GCCACCGCAT CAGAGACGTT	5100
GGACTTCAGC CATCCTCGGA GCTTCTCGTG TACTTCCTGG GCCTAGAACA AGAAGCTGGC	5160
CTAAGTAAGA CCTTTTCTGC CTCTCTAAGA GGAAAAATCA CTGGCACCAG TGGACACTTA	5220
GTGTGGTTTC TGACTGAGTC AGAGTACCAG GGCTCTGATC CAAGCCAGGC CCTGGACTGG	5280
ATGCCCTTGG ACAAGTCACT GTCTCTGGGT TCAAGGTCTC TGTGTCTTTG AAATAAGGGG	5340
TTGCCCCATG TGGGCTGTGT CTGTCCAAAC CTATTGAGGC AGGCTGGGAT GAGGGCAGGG	5400
CTCCTGGGCC CGGTACCTG TTGGGGTGTT GCAGTCTTGC CAGTACCAAT GGCCACACA	5460

- 180 -

GGCTCATAGG CCAGGACGAC CTTGCTCCAG TCCTTCACGT TATCTGCAGG GCAGAGATAC 5520
 AGATGGAGGG AAGGGTGAAC AAGAAAGAGC TCTCCAGCCA GGTCTCCGG AGTACGAAGA 5580
 ACGGTGGCCT ACTGCCCCCT AGTGGACATT GGGGG 5615

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 263 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Met	Gly	Gln	Thr	Ala	Leu	Ala	Gly	Gly	Ser	Ser	Ser	Thr	Pro	Thr	Pro
1				5					10					15	
Gln	Ala	Leu	Tyr	Pro	Asp	Leu	Ser	Cys	Pro	Glu	Gly	Leu	Glu	Glu	Leu
			20					25					30		
Leu	Ser	Ala	Pro	Pro	Pro	Asp	Leu	Gly	Ala	Gln	Arg	Arg	His	Gly	Trp
		35					40					45			
Asn	Pro	Lys	Asp	Cys	Ser	Glu	Asn	Ile	Glu	Val	Lys	Glu	Gly	Gly	Leu
		50					55				60				
Tyr	Phe	Glu	Arg	Arg	Pro	Val	Ala	Gln	Ser	Thr	Asp	Gly	Ala	Arg	Gly
65				70					75					80	
Lys	Arg	Gly	Tyr	Ser	Arg	Gly	Leu	His	Ala	Trp	Glu	Ile	Ser	Trp	Pro
			85					90					95		
Leu	Glu	Gln	Arg	Gly	Thr	His	Ala	Val	Val	Gly	Val	Ala	Thr	Ala	Leu
			100					105					110		
Ala	Pro	Leu	Gln	Thr	Asp	His	Tyr	Ala	Ala	Leu	Leu	Gly	Ser	Asn	Ser
			115					120					125		

- 181 -

Glu Ser Trp Gly Trp Asp Ile Gly Arg Gly Lys Leu Tyr His Gln Ser
 130 135 140

Lys Gly Pro Gly Ala Pro Gln Tyr Pro Ala Gly Thr Gln Gly Glu Gln
 145 150 155 160

Leu Glu Val Pro Glu Arg Leu Leu Val Val Leu Asp Met Glu Glu Gly
 165 170 175

Thr Leu Gly Tyr Ala Ile Gly Gly Thr Tyr Leu Gly Pro Ala Phe Arg
 180 185 190

Gly Leu Lys Gly Arg Thr Leu Tyr Pro Ala Val Ser Ala Val Trp Gly
 195 200 205

Gln Cys Gln Val Arg Ile Arg Tyr Leu Gly Glu Arg Arg Ala Glu Pro
 210 215 220

His Ser Leu Leu His Leu Ser Arg Leu Cys Val Arg His Asn Leu Gly
 225 230 235 240

Asp Thr Arg Leu Gly Gln Val Ser Ala Leu Pro Leu Pro Pro Ala Met
 245 250 255

Lys Arg Tyr Leu Leu Tyr Gln
 260

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

AGCTAGATCTGGACCCTACAATGGCAGC

28

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- 182 -

- (A) LENGTH: base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

AGCTAGATCT GCCATCCTAC TCGAGGGGCC AGCTGG

36

- 183 -

CLAIMS:

1. A nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a protein or a derivative, homologue, analogue or mimetic thereof or a nucleotide sequence capable of hybridizing thereto under low stringency conditions at 42°C wherein said protein comprises a SOCS box in its C-terminal region.
2. A nucleic acid molecule according to claim 1 wherein the protein further comprises a protein:molecule interacting region.
3. A nucleic acid molecule according to claim 1 wherein the protein:molecule interacting region is located in a region N-terminal of the SOCS box.
4. A nucleic acid molecule according to claim 2 or 3 wherein the protein:molecule interacting region is a protein:DNA binding region or a protein:protein binding region.
5. A nucleic acid molecule according to claim 4 wherein the protein:molecule interacting region is one or more of an SH2 domain, WD-40 repeats or ankyrin repeats.
6. A nucleic acid molecule according to any one of claims 1-5 wherein the SOCS box comprises the amino acid sequence:

$$X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14} X_{15} X_{16} [X_i]_n X_{17} X_{18} X_{19} X_{20} \\ X_{21} X_{22} X_{23} [X_j]_n X_{24} X_{25} X_{26} X_{27} X_{28}$$

wherein:

- X_1 is L, I, V, M, A or P;
- X_2 is any amino acid residue;
- X_3 is P, T or S;
- X_4 is L, I, V, M, A or P;
- X_5 is any amino acid;
- X_6 is any amino acid;

- 184 -

X_7 is L, I, V, M, A, F, Y or W;

X_8 is C, T or S;

X_9 is R, K or H;

X_{10} is any amino acid;

X_{11} is any amino acid;

X_{12} is L, I, V, M, A or P;

X_{13} is any amino acid;

X_{14} is any amino acid;

X_{15} is any amino acid;

X_{16} is L, I, V, M, A, P, G, C, T or S;

$[X_i]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;

X_{17} is L, I, V, M, A or P;

X_{18} is any amino acid;

X_{19} is any amino acid;

X_{20} L, I, V, M, A or P;

X_{21} is P;

X_{22} is L, I, V, M, A, P or G;

X_{23} is P or N;

$[X_j]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_j may comprise the same or different amino acids selected from any amino acid residue;

X_{24} is L, I, V, M, A or P;

X_{25} is any amino acid;

X_{26} is any amino acid;

X_{27} is Y or F; and

X_{28} is L, I, V, M, A or P.

7. A nucleic acid molecule according to claim 6 wherein the protein modulates signal transduction.

- 185 -

8. A nucleic acid molecule according to claim 7 wherein the signal transduction is modulated by a cytokine or a hormone, a microbe or a microbial product, a parasite, an antigen or other effector molecule.

9. A nucleic acid molecule according to claim 8 wherein the protein modulates cytokine-mediated signal transduction.

10. A nucleic acid molecule according to claim 9 wherein the signal transduction is mediated by one or more of the cytokines EPO, TPO, G-CSF, GM-CSF, IL-3, IL-2, IL-4, IL-7, IL-13, IL-6, LIF, IL-12, IFN γ , TNF α , IL-1 and/or M-CSF.

11. A nucleic acid molecule according to claim 10 wherein the signal transduction is mediated by one or more of IL-6, LIF, OSM, IFN- γ and/or thrombopoietin.

12. A nucleic acid molecule according to claim 11 wherein the signal transduction is mediated by IL-6.

13. A nucleic acid molecule according to claim 1 wherein the nucleotide sequence encodes an amino acid sequence substantially as set forth in SEQ ID NO. 4, SEQ ID NO. 6, SEQ ID NO. 8, SEQ ID NO. 10, SEQ ID NO. 12, SEQ ID NO. 14, SEQ ID NO. 18, SEQ ID NO. 21, SEQ ID NO. 25, SEQ ID NO. 29, SEQ ID NO. 36, SEQ ID NO. 41, SEQ ID NO. 44, SEQ ID NO. 46 or SEQ ID NO. 48 or an amino acid sequence having at least about 15% similarity to all or part of the listed sequences or a nucleotide sequence which hybridizes to the nucleic acid molecule under low stringency conditions at 42°C.

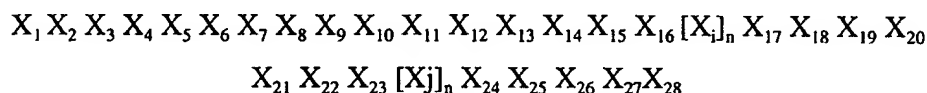
14. A nucleic acid molecule according to claim 1 wherein the nucleotide sequence is substantially as set forth in SEQ ID NO. 3, SEQ ID NO. 5, SEQ ID NO. 7, SEQ ID NO. 9, SEQ ID NO. 11, SEQ ID NO. 13, SEQ ID NO. 15, SEQ ID NO. 16, SEQ ID NO. 17, SEQ ID NO. 20, SEQ ID NO. 22, SEQ ID NO. 23, SEQ ID NO. 24, SEQ ID NO. 26, SEQ ID NO. 27, SEQ ID NO. 28, SEQ ID NO. 30, SEQ ID NO. 31, SEQ ID NO. 32, SEQ ID NO. 33, SEQ ID NO. 34, SEQ ID NO. 35, SEQ ID NO. 37, SEQ ID NO. 38, SEQ ID NO. 39, SEQ ID NO. 40, SEQ

- 186 -

ID NO. 42, SEQ ID NO. 43, SEQ ID NO. 45 or SEQ ID NO. 47 or a nucleotide sequence having at least 15% similarity to all or a part of the listed sequences or a nucleotide sequence capable of hybridizing to the listed sequences under low stringency conditions at 42°C.

15. A nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a protein or a derivative, homologue, analogue or mimetic thereof or a nucleotide sequence capable of hybridizing thereto under low stringency conditions at 42°C wherein said protein exhibits the following characteristics:

- (i) comprises a SOCS box in its C-terminal region wherein said SOCS box comprises the amino acid sequence:



wherein:

- X_1 is L, I, V, M, A or P;
- X_2 is any amino acid residue;
- X_3 is P, T or S;
- X_4 is L, I, V, M, A or P;
- X_5 is any amino acid;
- X_6 is any amino acid;
- X_7 is L, I, V, M, A, F, Y or W;
- X_8 is C, T or S;
- X_9 is R, K or H;
- X_{10} is any amino acid;
- X_{11} is any amino acid;
- X_{12} is L, I, V, M, A or P;
- X_{13} is any amino acid;
- X_{14} is any amino acid;
- X_{15} is any amino acid;
- X_{16} is L, I, V, M, A, P, G, C, T or S;

- 187 -

$[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;

X_{17} is L, I, V, M, A or P;

X_{18} is any amino acid;

X_{19} is any amino acid;

X_{20} L, I, V, M, A or P;

X_{21} is P;

X_{22} is L, I, V, M, A, P or G;

X_{23} is P or N;

$[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_j may comprise the same or different amino acids selected from any amino acid residue;

X_{24} is L, I, V, M, A or P;

X_{25} is any amino acid;

X_{26} is any amino acid;

X_{27} is Y or F;

X_{28} is L, I, V, M, A or P; and

- (ii) comprises at least one of an SH2 domain, WD-40 repeats and/or ankyrin repeats or other protein:molecule interacting domain in a region N-terminal of the SOCS box; and
- (iii) modulates signal transduction.

16. An isolated protein or a derivative, homologue or mimetic thereof comprising a SOCS box in its C-terminal region.

17. An isolated protein according to claim 16 wherein the protein further comprises a protein:molecule interacting region.

18. An isolated protein according to claim 17 wherein the protein:molecule interacting region

is located in a region N-terminal of the SOCS box.

19. An isolated protein according to claim 16 or 17 wherein the protein:molecule interacting region is a protein:DNA binding region or a protein:protein binding region.
20. An isolated protein according to claim 19 wherein the protein:molecule interacting region is one or more of an SH2 domain, WD-40 repeats or ankyrin repeats.
21. An isolated protein according to any one of claims 16-20 wherein the SOCS box comprises the amino acid sequence:

[illegible]

wherein:

X_1 is L, I, V, M, A or P;

X_2 is any amino acid residue;

X_3 is P, T or S;

X_4 is L, I, V, M, A or P;

X_5 is any amino acid;

X_6 is any amino acid;

X_7 is L, I, V, M, A, F, Y or W;

X_8 is C, T or S;

X_9 is R, K or H;

X_{10} is any amino acid;

X_{11} is any amino acid;

X_{12} is L, I, V, M, A or P;

X_{13} is any amino acid;

X_{14} is any amino acid;

X_{15} is any amino acid;

X_{16} is L, I, V, M, A, P, G, C, T or S;

$[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids

- 189 -

and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;

X_{17} is L, I, V, M, A or P;

X_{18} is any amino acid;

X_{19} is any amino acid;

X_{20} L, I, V, M, A or P;

X_{21} is P;

X_{22} is L, I, V, M, A, P or G;

X_{23} is P or N;

$[X_j]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_j may comprise the same or different amino acids selected from any amino acid residue;

X_{24} is L, I, V, M, A or P;

X_{25} is any amino acid;

X_{26} is any amino acid;

X_{27} is Y or F; and

X_{28} is L, I, V, M, A or P.

22. An isolated protein according to claim 21 wherein the protein modulates signal transduction.

23. An isolated protein according to claim 22 wherein the signal transduction is modulated by a cytokine or other endogenous molecule, a hormone, a microbe or a microbial product, a parasite, an antigen or other effector molecule.

24. An isolated protein according to claim 23 wherein the protein modulates cytokine-mediated signal transduction.

25. An isolated protein according to claim 24 wherein the signal transduction is mediated by one or more of the cytokines EPO, TPO, G-CSF, GM-CSF, IL-3, IL-2, IL-4, IL-7, IL-13, IL-6, LIF, IL-12, IFN γ , TNF α , IL-1 and/or M-CSF.

- 190 -

26. An isolated protein according to claim 25 wherein the signal transduction is mediated by one or more of IL-6, LIF, OSM, IFN- γ and/or thrombopoietin.

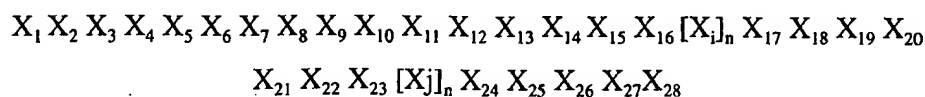
27. An isolated protein according to claim 26 wherein the signal transduction is mediated by IL-6.

28. An isolated protein according to claim 16 wherein said protein comprises an amino acid sequence substantially as set forth in SEQ ID NO. 4, SEQ ID NO. 6, SEQ ID NO. 8, SEQ ID NO. 10, SEQ ID NO. 12, SEQ ID NO. 14, SEQ ID NO. 18, SEQ ID NO. 21, SEQ ID NO. 25, SEQ ID NO. 29, SEQ ID NO. 36, SEQ ID NO. 41, SEQ ID NO. 44, SEQ ID NO. 46 or SEQ ID NO. 48 or an amino acid sequence having at least about 15% similarity to all or part of the listed sequences.

29. An isolated protein according to claim 16 wherein the said protein is encoded by a nucleotide sequence substantially as set forth in SEQ ID NO. 3, SEQ ID NO. 5, SEQ ID NO. 7, SEQ ID NO. 9, SEQ ID NO. 11, SEQ ID NO. 13, SEQ ID NO. 15, SEQ ID NO. 16, SEQ ID NO. 17, SEQ ID NO. 20, SEQ ID NO. 22, SEQ ID NO. 23, SEQ ID NO. 24, SEQ ID NO. 26, SEQ ID NO. 27, SEQ ID NO. 28, SEQ ID NO. 30, SEQ ID NO. 31, SEQ ID NO. 32, SEQ ID NO. 33, SEQ ID NO. 34, SEQ ID NO. 35, SEQ ID NO. 37, SEQ ID NO. 38, SEQ ID NO. 39, SEQ ID NO. 40, SEQ ID NO. 42, SEQ ID NO. 43, SEQ ID NO. 45 or SEQ ID NO. 47 or a nucleotide sequence having at least 15% similarity to all or a part of the listed sequences or a nucleotide sequence capable of hybridizing to the listed sequences under low stringency conditions at 42°C.

30. An isolated protein or a derivative, homologue, analogue or mimetic thereof having the following characteristics:

- (i) comprises a SOCS box in its C-terminal region wherein said SOCS box comprises the amino acid sequence:



- 191 -

wherein:

- X_1 is L, I, V, M, A or P;
- X_2 is any amino acid residue;
- X_3 is P, T or S;
- X_4 is L, I, V, M, A or P;
- X_5 is any amino acid;
- X_6 is any amino acid;
- X_7 is L, I, V, M, A, F, Y or W;
- X_8 is C, T or S;
- X_9 is R, K or H;
- X_{10} is any amino acid;
- X_{11} is any amino acid;
- X_{12} is L, I, V, M, A or P;
- X_{13} is any amino acid;
- X_{14} is any amino acid;
- X_{15} is any amino acid;
- X_{16} is L, I, V, M, A, P, G, C, T or S;
- $[X_i]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;
- X_{17} is L, I, V, M, A or P;
- X_{18} is any amino acid;
- X_{19} is any amino acid;
- X_{20} L, I, V, M, A or P;
- X_{21} is P;
- X_{22} is L, I, V, M, A, P or G;
- X_{23} is P or N;
- $[X_j]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_j may comprise the same or different amino acids selected from any amino acid residue;
- X_{24} is L, I, V, M, A or P;
- X_{25} is any amino acid;

- 192 -

X₂₆ is any amino acid;

X₂₇ is Y or F;

X₂₈ is L, I, V, M, A or P; and

- (ii) comprises at least one of an SH2 domain, WD-40 repeats and/or ankyrin repeats or other protein:molecule interacting domain in a region N-terminal of the SOCS box; and
 - (iii) modulates signal transduction.
31. A method of modulating levels of a SOCS protein in a cell said method comprising contacting a cell containing a SOCS gene with an effective amount of a modulator of SOCS gene expression or SOCS protein activity for a time and under conditions sufficient to modulate levels of said SOCS protein.
32. A method of modulating signal transduction in a cell containing a SOCS gene comprising contacting said cell with an effective amount of a modulator of SOCS gene expression or SOCS protein activity for a time sufficient to modulate signal transduction.
33. A method of influencing interaction between cells wherein at least one cell carries a SOCS gene, said method comprising contacting the cell carrying the SOCS gene with an effective amount of a modulator of SOCS gene expression or SOCS protein activity for a time sufficient to modulate signal transduction.
34. A method according to any one of claims 31-33 wherein signal transduction is mediated by a cytokine, a hormone, a microbe or a microbial product, a parasite, an antigen or other effector molecule.
35. A method according to claim 34 wherein the cytokine is one or more of EPO, TPO, G-CSF, GM-CSF, IL-3, IL-2, IL-4, IL-7, IL-13, IL-6, LIF, IL-12, IFN γ , TNF α , IL-1 and/or M-CSF.

- 193 -

36. A method according to claim 35 wherein the cytokine is one or more of IL-6, LIF, OSM, IFN- γ and/or thrombopoietin.
37. A method according to claim 36 wherein the cytokine is IL-6.
38. A method according to any one of claims 31-37 wherein the SOCS gene encodes a protein having a SOCS box comprising the amino acid sequence:

$$X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10} X_{11} X_{12} X_{13} X_{14} X_{15} X_{16} [X]_n X_{17} X_{18} X_{19} X_{20} \\ X_{21} X_{22} X_{23} [X]_n X_{24} X_{25} X_{26} X_{27} X_{28}$$

wherein:

- X_1 is L, I, V, M, A or P;
- X_2 is any amino acid residue;
- X_3 is P, T or S;
- X_4 is L, I, V, M, A or P;
- X_5 is any amino acid;
- X_6 is any amino acid;
- X_7 is L, I, V, M, A, F, Y or W;
- X_8 is C, T or S;
- X_9 is R, K or H;
- X_{10} is any amino acid;
- X_{11} is any amino acid;
- X_{12} is L, I, V, M, A or P;
- X_{13} is any amino acid;
- X_{14} is any amino acid;
- X_{15} is any amino acid;
- X_{16} is L, I, V, M, A, P, G, C, T or S;
- $[X]_n$ is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_i may comprise the same or different amino acids selected from any amino acid residue;
- X_{17} is L, I, V, M, A or P;
- X_{18} is any amino acid;

- 194 -

X₁₉ is any amino acid;

X₂₀ L, I, V, M, A or P;

X₂₁ is P;

X₂₂ is L, I, V, M, A, P or G;

X₂₃ is P or N;

[X]_n is a sequence of n amino acids wherein n is from 1 to 50 amino acids and wherein the sequence X_j may comprise the same or different amino acids selected from any amino acid residue;

X₂₄ is L, I, V, M, A or P;

X₂₅ is any amino acid;

X₂₆ is any amino acid;

X₂₇ is Y or F; and

X₂₈ is L, I, V, M, A or P.

39. A method according to claim 38 wherein the SOCS gene comprises a nucleotide sequence selected from SEQ ID NO. 3, SEQ ID NO. 5, SEQ ID NO. 7, SEQ ID NO. 9, SEQ ID NO. 11, SEQ ID NO. 13, SEQ ID NO. 15, SEQ ID NO. 16, SEQ ID NO. 17, SEQ ID NO. 20, SEQ ID NO. 22, SEQ ID NO. 23, SEQ ID NO. 24, SEQ ID NO. 26, SEQ ID NO. 27, SEQ ID NO. 28, SEQ ID NO. 30, SEQ ID NO. 31, SEQ ID NO. 32, SEQ ID NO. 33, SEQ ID NO. 34, SEQ ID NO. 35, SEQ ID NO. 37, SEQ ID NO. 38, SEQ ID NO. 39, SEQ ID NO. 40, SEQ ID NO. 42, SEQ ID NO. 43, SEQ ID NO. 45 or SEQ ID NO. 47.

40. A method according to claim 38 wherein the SOCS gene encodes a protein comprising an amino acid sequence substantially as set forth in SEQ ID NO. 4, SEQ ID NO. 6, SEQ ID NO. 8, SEQ ID NO. 10, SEQ ID NO. 12, SEQ ID NO. 14, SEQ ID NO. 18, SEQ ID NO. 21, SEQ ID NO. 25, SEQ ID NO. 29, SEQ ID NO. 36, SEQ ID NO. 41, SEQ ID NO. 44, SEQ ID NO. 46 or SEQ ID NO. 48.

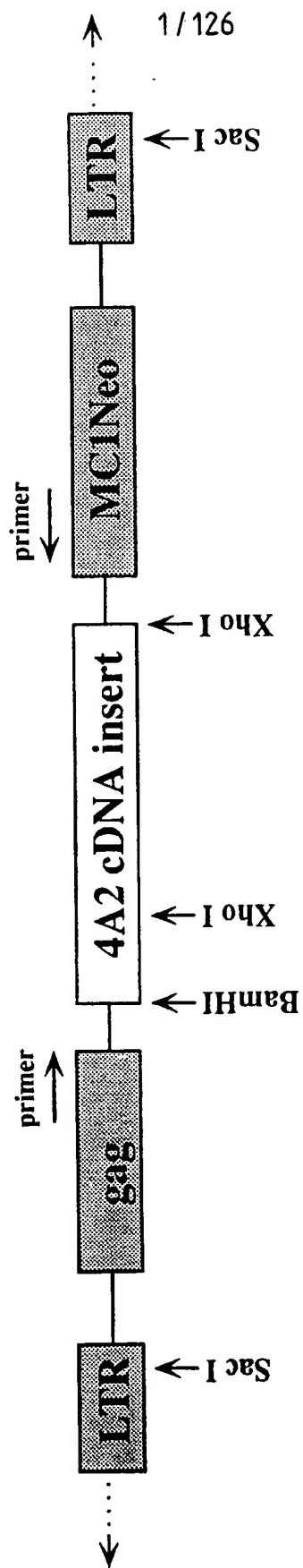


FIGURE 1

2/126

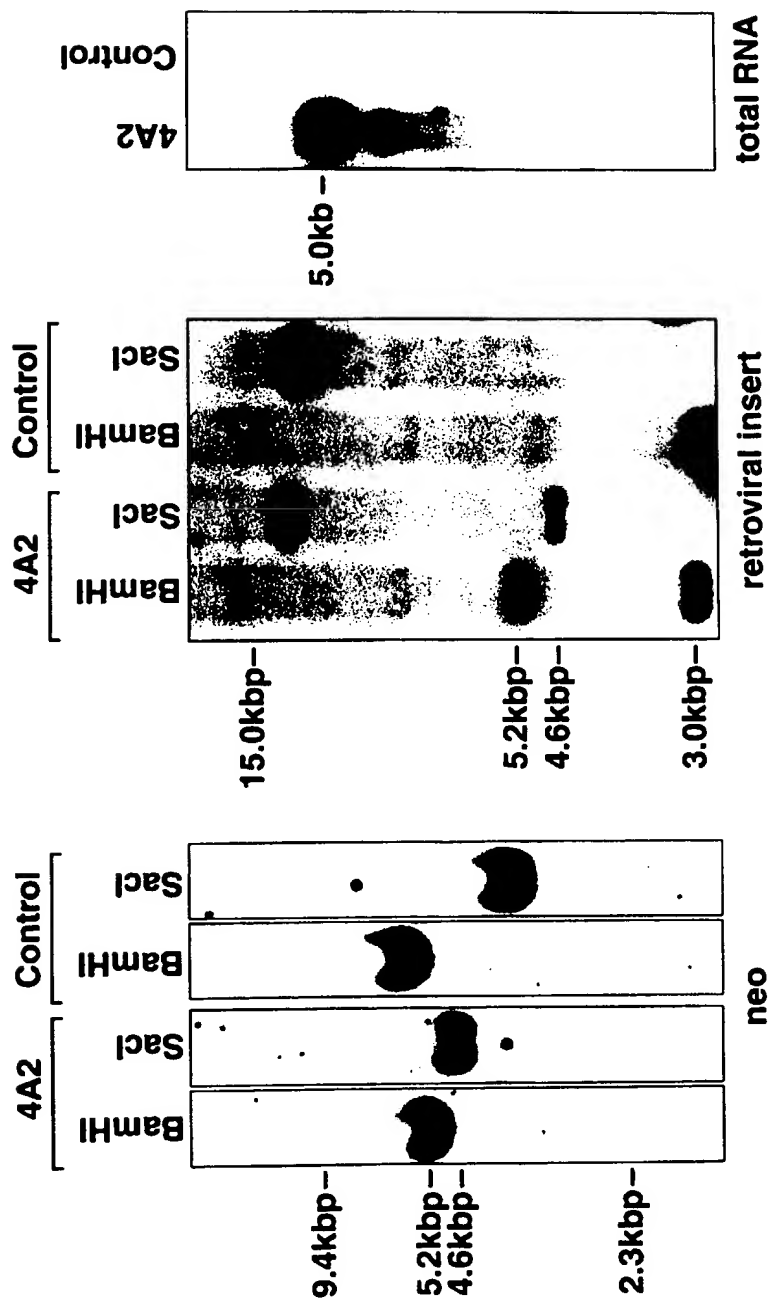


FIG 2

3/126

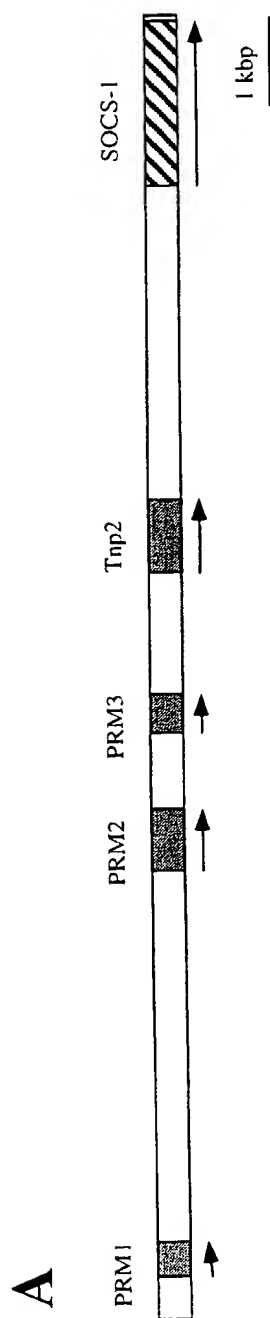


FIG 3A

4/126

-159 cgaggctcaagctccgggaggattctgCGTgCCgctctcg
-120 ctcccttggggctctgttggccggcctgtgccaccggagcccggtcactgectctgtct
-60 ccccatcagcgcagcccgagcgtatggccaccctccagctggccccctcgagtagg

1 M V A R N Q V A A D N A I S P A A E P R
1 ATGGTAGCACGCAACCAGGTGGCAGCCGACAATGCGATCTCCCCGGCAGCAGAGCCCCGA

21 R R S E P S S S S S S S S S P A A P V R P
61 CGGCGGTcAGAGCCCTCCTCGTCCTCGTCTTCGTCTCGCCAGCGGCCCCCGTGCCTCCC

41 R P C P A V P A P A P G D T H F R T F R
121 CGGCCCTGCCCCGGCGGTCCCAGCCCCAGCCCTGGCGACACTCACTTCCGCACCTTCCGC

61 S H S D Y R R I T R T S A L L D A C G F
181 TCCCACTCCGATTACCGGCGCATCACGCGGACCAGCGGCTCCTGGACGCTTGGCGCTTC

81 Y W G P L S V H G A H E R L R A E P V G
241 TATTGGGGACCCCTGAGCGTGCACGGGGCGCACGAGCGGCTGCGTGCCGAGCCCGTGGGG

101 T F L V R D S R Q R N C F F A L S V K M
301 ACCTTCTTGGTGCGCGACAGTCGTCAACGGAAGTCTTCTTCGCGCTCAGCGTGAAGATG

121 A S G P T S I R V H F Q A G R F H L D G
361 GCTTCGGGCCCCACGAGCATCCGCGTGCCTTCCAGCGCGCGCTTCCACTTGGACGGC

141 S R E T F D C L F E L L E H Y V A A P R
421 AGCCGCGAGACCTTCGACTGCCTTTTCGAGCTGCTGGAGCAcTACGTGGCGGCGCCGCGC

161 R M L G A P L R Q R R V R P L Q E L C R
481 CGCATGTGGGGGGCCcGCTGCGCCAGCGCCGCTGCGGCCGCTGCAGGAGCTGTGTGCG

181 Q R I V A A V G R E N L A R I P L N P V
541 CAGCGCATCGTGGCCGCGTGGTTCGCGAGAACCTGGCGCGCATCCcTCTTAACCCGGTA

201 L R D Y L S S F P F Q I *
601 CTCCGTGACTACCTGAGTTCCTTCCCCTTCCAGATCtgaccggctgCCgctgtgCCgCag

661 cattaagtgggggcgccttattattttcttattattaattattattatttttctggaacca
721 cgtgggagccctccccgcctgggtcggagggtggttgaggagggtgagatgcctccca
781 cttctggctggagacctcatcccacctctcaggggtgggggtgctccctcctggtgctc
841 cctccgggtccccctgggttagcagcttgtgtctggggccaggacctgaattccactc
901 ctacctctccatgtttacatattccagtatctttgcacaaaccaggggtcggggagggt
961 ctctggcttcatttttctgctgtgcagaatattctattttatatttttacagccagttta
1021 ggtataaaactttattatgaaagtttttttttaaaagaaaaaaaaaaaaaaaaaaaaa

FIG 3B

5/126

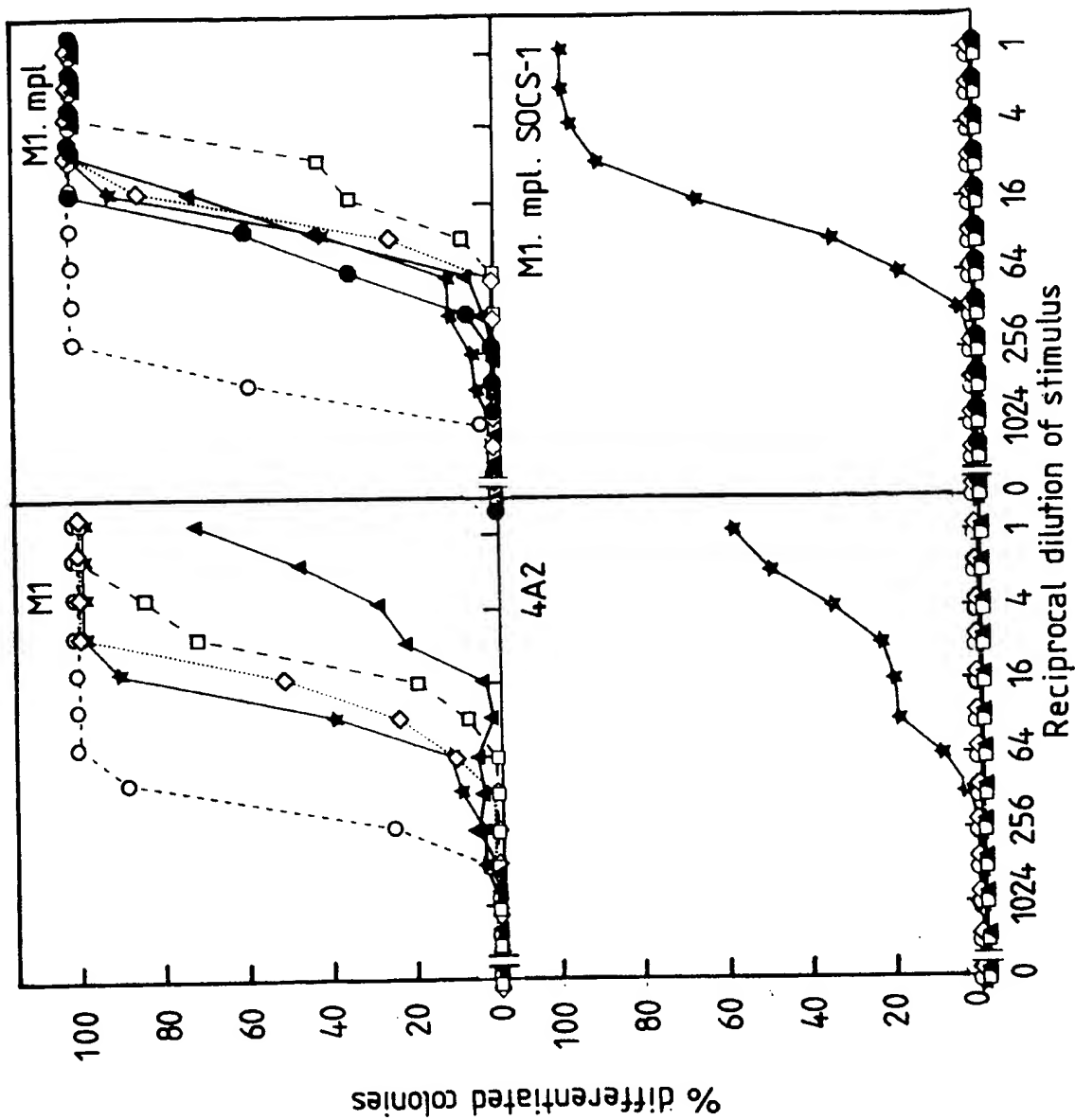


FIG 4

6/126

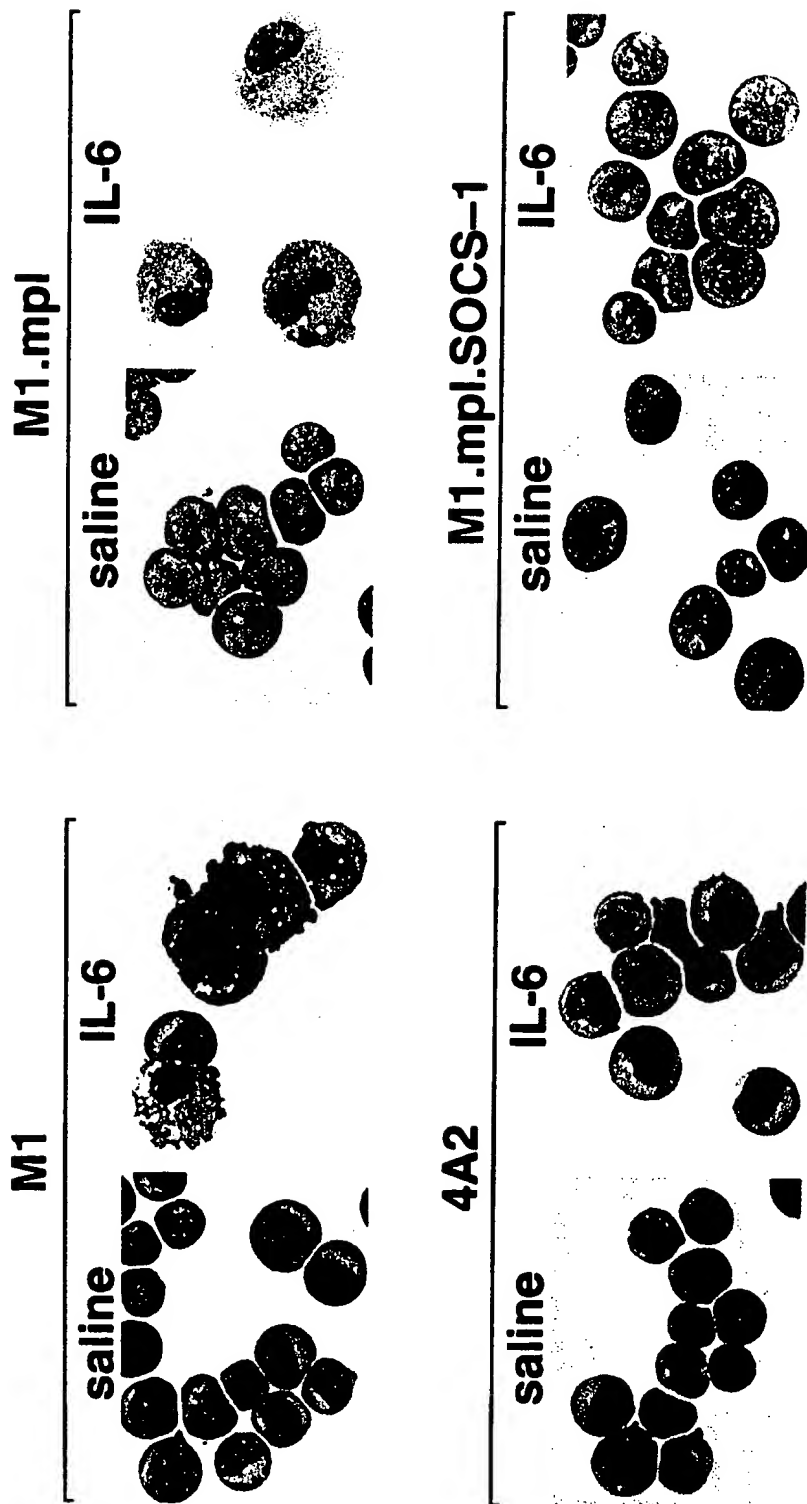


FIG 5

7/126

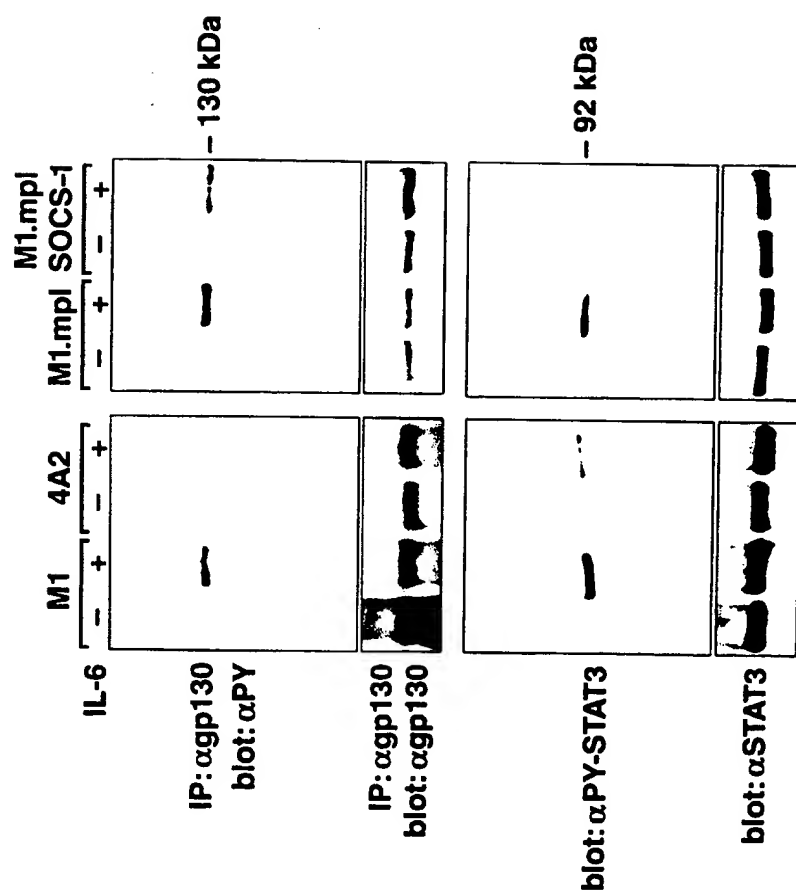
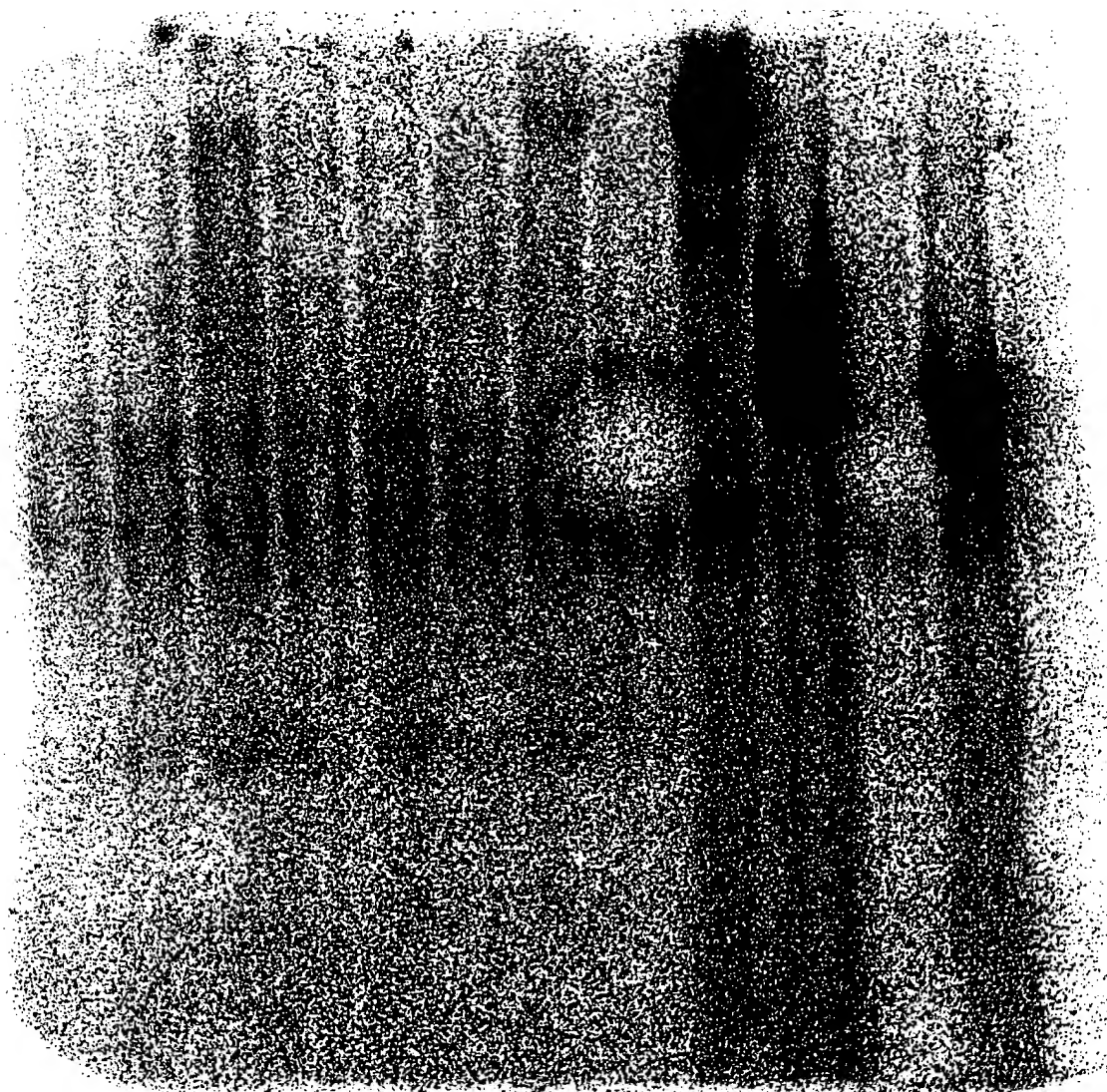


FIG 6

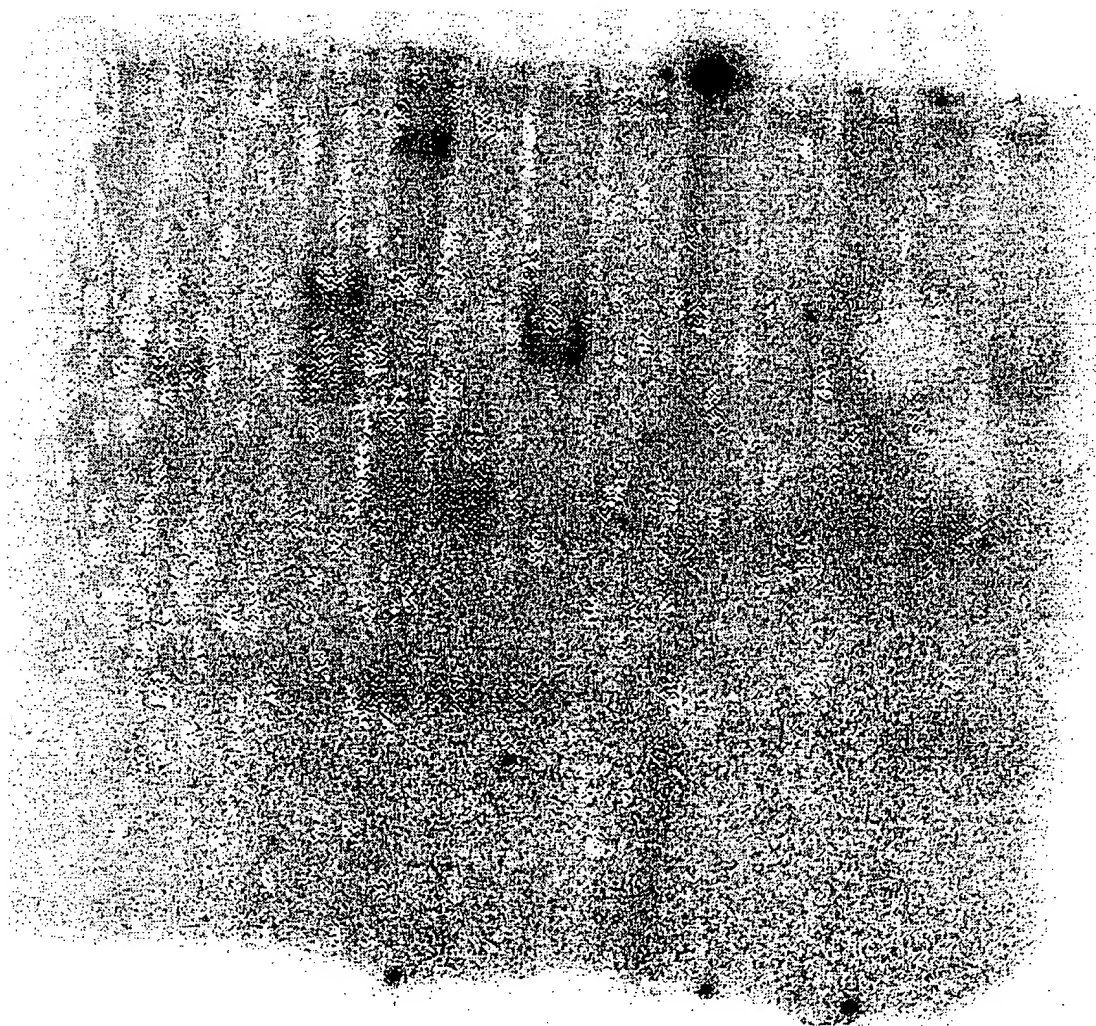
8/126

FIG 7A



9/126

FIG 7B



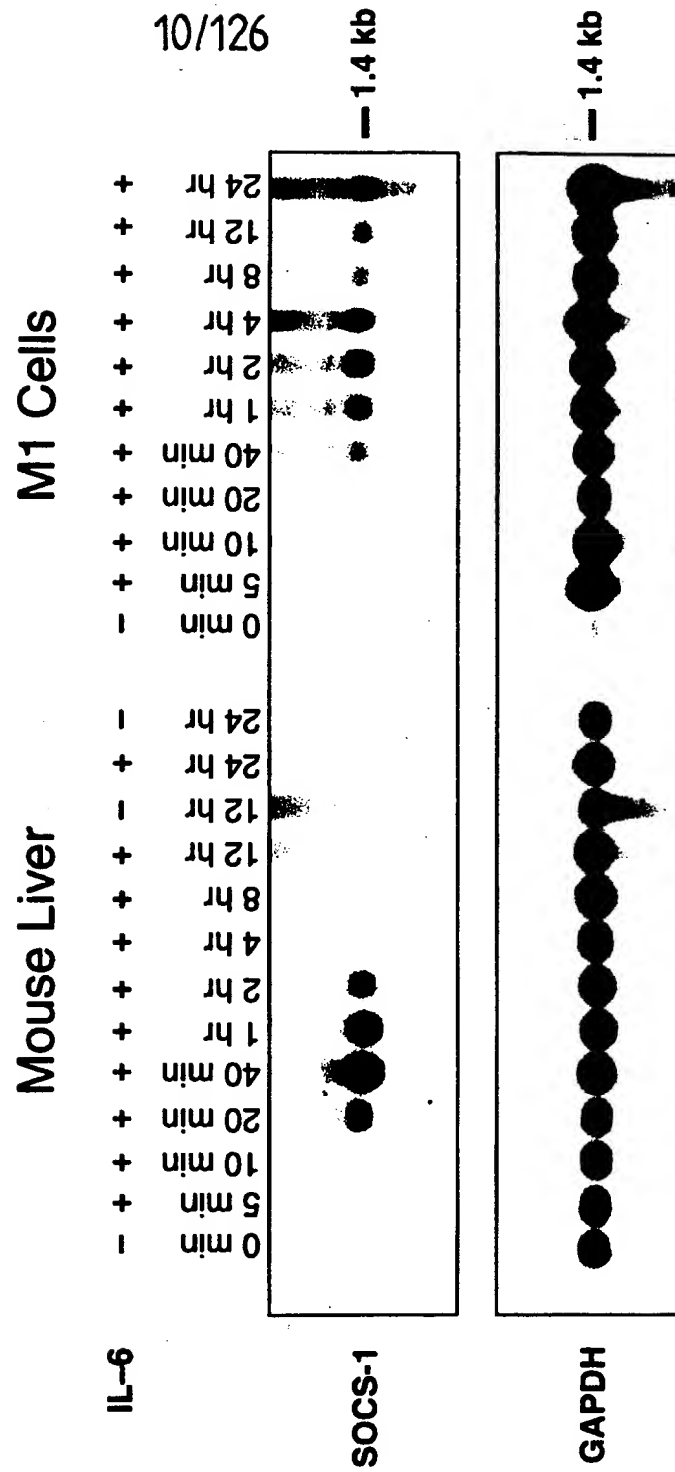


FIG 8

11/126

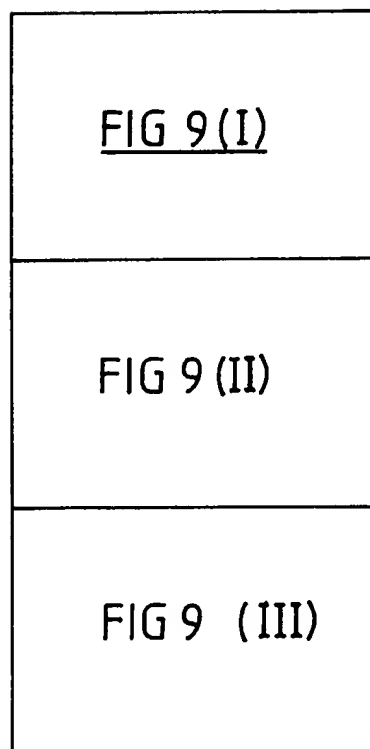


FIG 9

hs SOCS-1	(98)	V	G	T	F	L	V	R	D	S	R	Q	R	N	C	F	F	A	L	S	V	K	M	A	S	E	P	T	S	I	R	V	H	F	Q	A	G	R	F	H	I	D	G	S	R	(141)
rr SOCS-1	(99)	V	G	T	F	L	V	R	D	S	R	Q	R	N	C	F	F	A	L	S	V	K	M	A	S	E	P	T	S	I	R	V	H	F	Q	A	G	R	F	H	I	D	G	N	R	(142)
mm SOCS-1	(99)	V	G	T	F	L	V	R	D	S	R	Q	R	N	C	F	F	A	L	S	V	K	M	A	S	E	P	T	S	I	R	V	H	F	Q	A	G	R	F	H	I	D	G	S	R	(142)
mm SOCS-2	(67)	E	G	T	F	L	V	R	D	S	S	H	S	D	V	L	L	T	S	N	K	T	S	A	E	P	T	N	L	R	I	E	Y	Q	D	G	K	F	R	I	D	S	I	I	C	V	K	S	K	L	(116)	
mm SOCS-3	(65)	A	G	T	F	L	V	R	D	S	S	D	Q	R	H	F	F	L	S	V	K	T	Q	S	G	T	K	N	L	R	I	Q	C	E	G	G	S	F	S	L	Q	S	D	P	R	S	T	Q	P	V	(117)	
mm CIS	(101)	E	G	T	F	L	V	R	D	S	T	H	P	S	Y	X	E	T	L	S	V	R	T	T	R	Q	P	T	N	V	R	I	E	Y	A	D	S	S	E	R	L	D	S	N	C	L	S	R	P	R	I	(150)

hs SOCS-1	(142)	E	S	F	D	C	L	F	E	L	E	H	V	V	A	A	P	R	R	M	L	G	A	P	(165)
rr SOCS-1	(143)	E	T	F	D	C	L	F	E	L	E	H	V	V	A	A	P	R	R	M	L	G	A	P	(166)
mm SOCS-1	(143)	E	T	F	D	C	L	F	E	L	E	H	V	V	A	A	P	R	R	M	L	G	A	P	(166)
mm SOCS-2	(117)	K	Q	E	D	S	V	V	H	L	I	D	Y	V	Q	M	C	K	D	K	R	T	G	P	(140)	
mm SOCS-3	(115)	P	R	F	D	C	V	L	K	L	V	H	H	V	M	P	P	P	G	T	P	S	F	S	L	P	P	T	E	P	S	S	E	V	P	E	Q	P	P	A	Q	A	L	P	G	S	T	P	K	R	A	(164)	
mm CIS	(151)	L	A	E	P	D	V	V	S	V	X	Q	H	V	V	A	S	C	A	A	D	T	R	S	D	S	P	D	P	A	P	T	P	A	L	P	M	S	K	Q	D	A	P	S	D	S	V	L	P	I	P	(200)	

FIG 9(II)

15/126

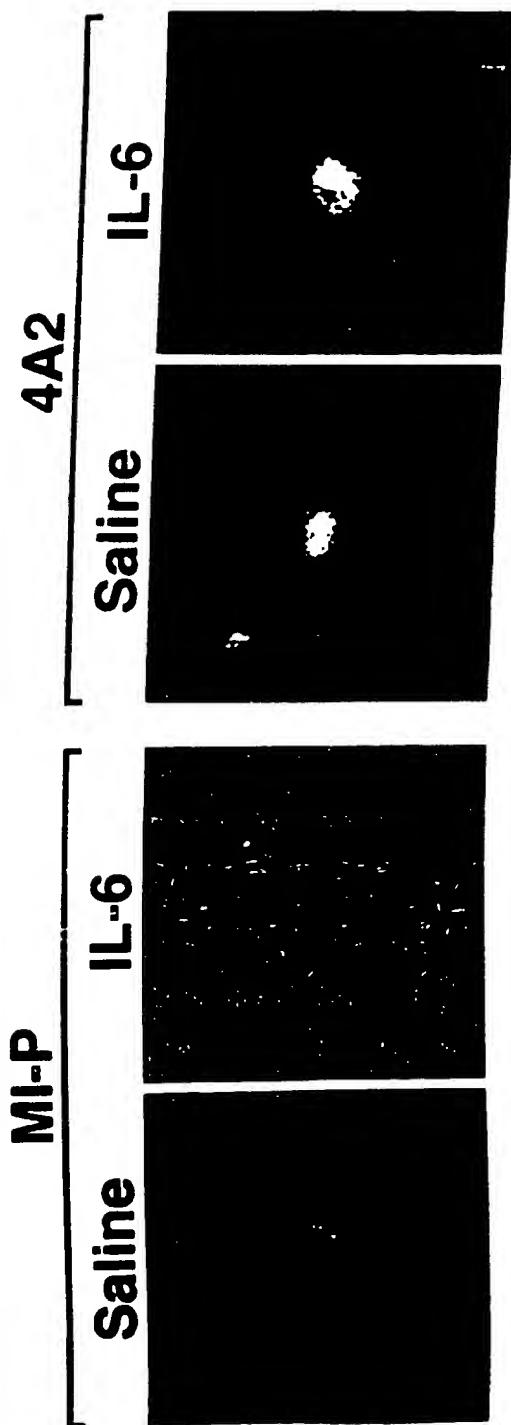
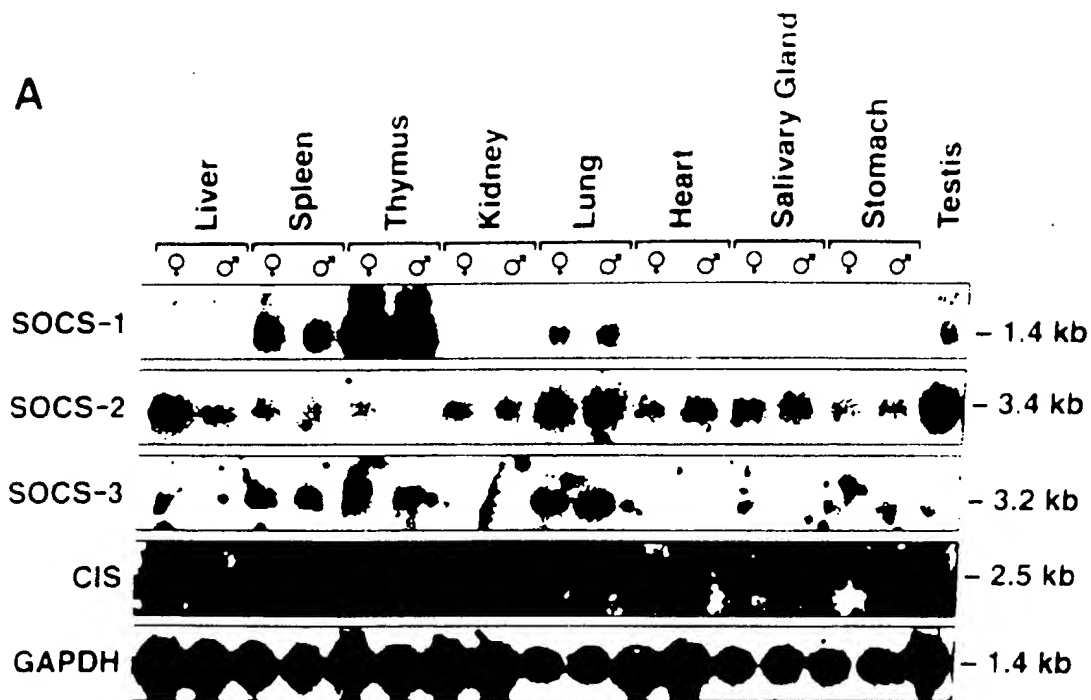


FIG 10

16/126

FIG 11A

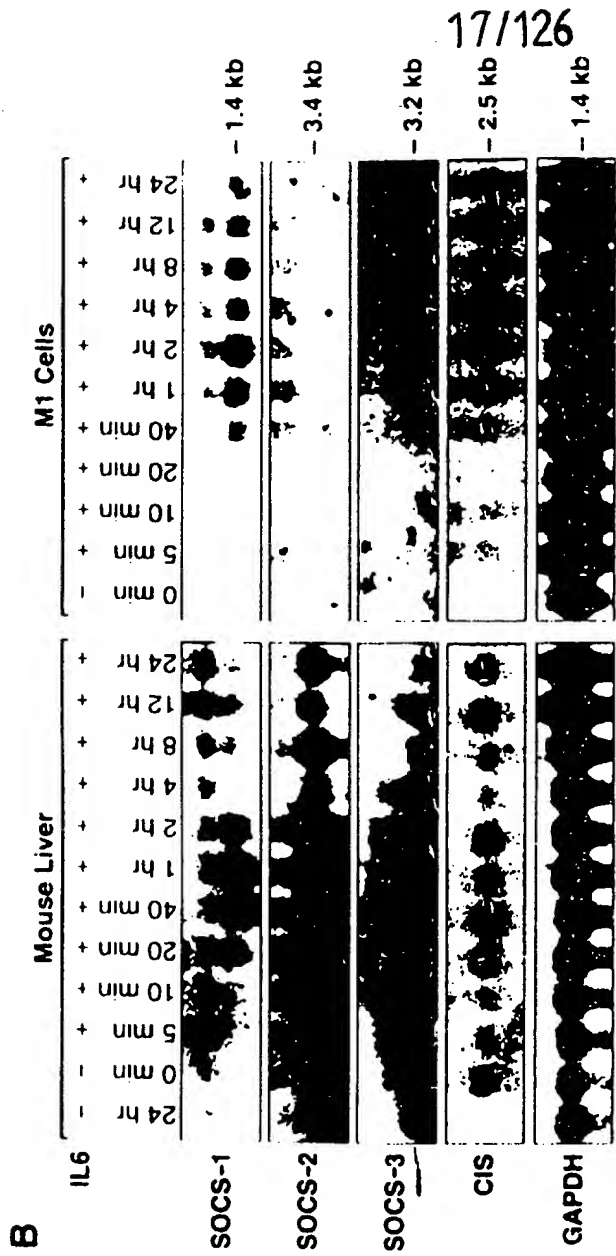


FIG 11B

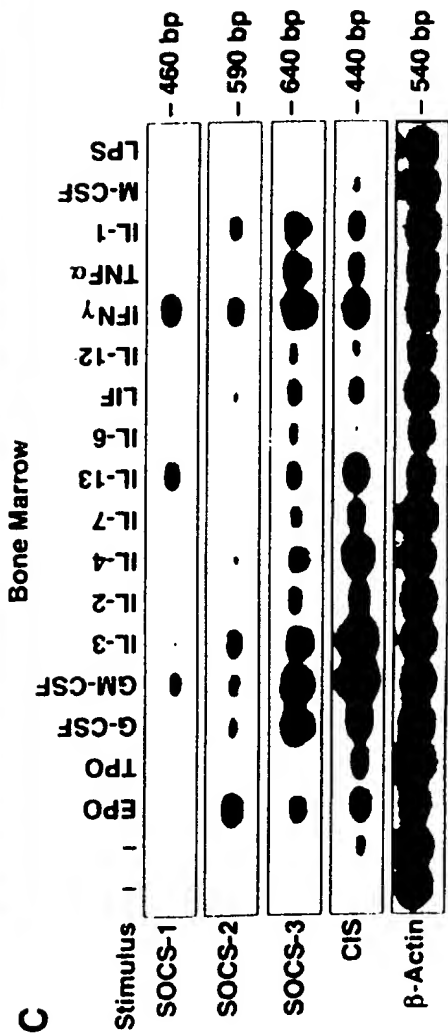


FIG 11C

18/126

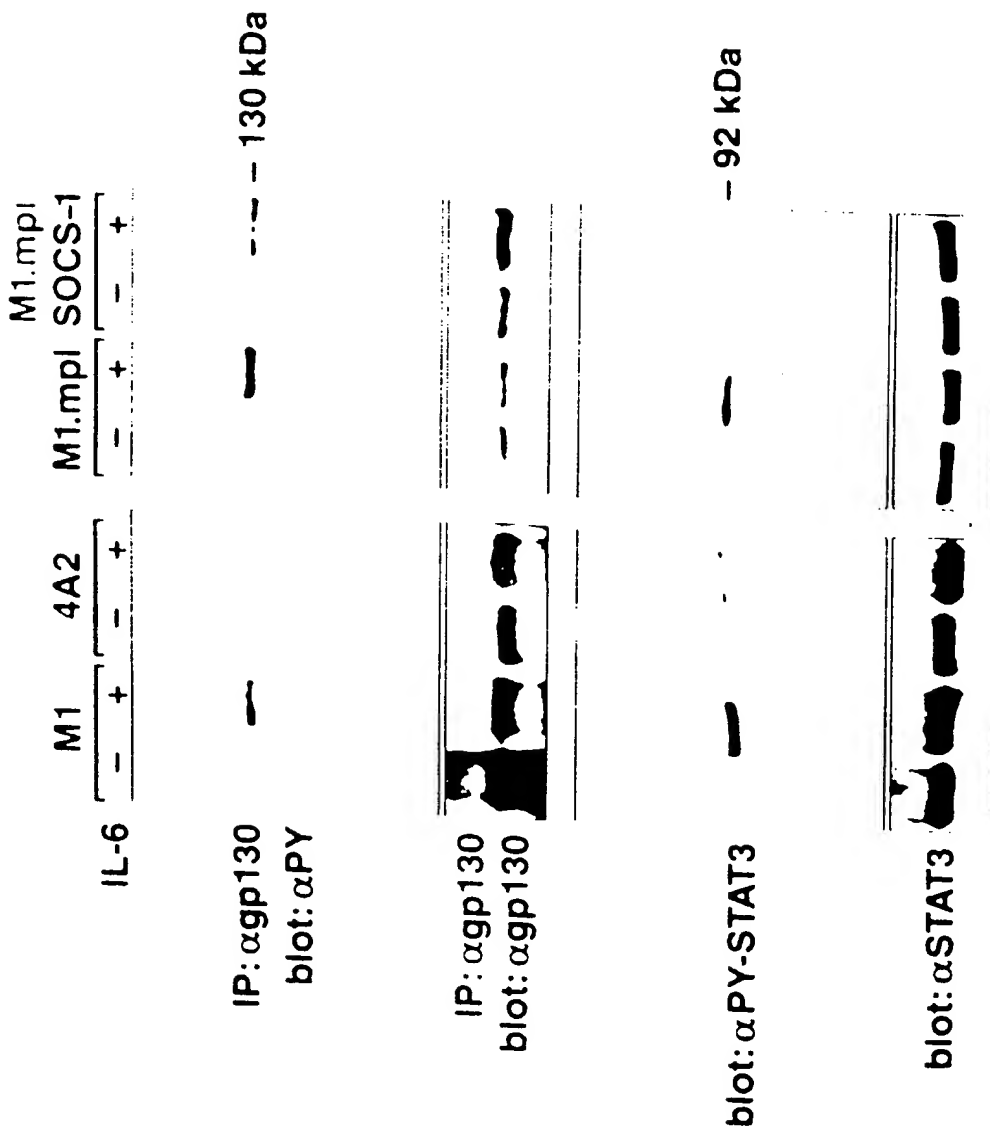


FIG 12A

19/126

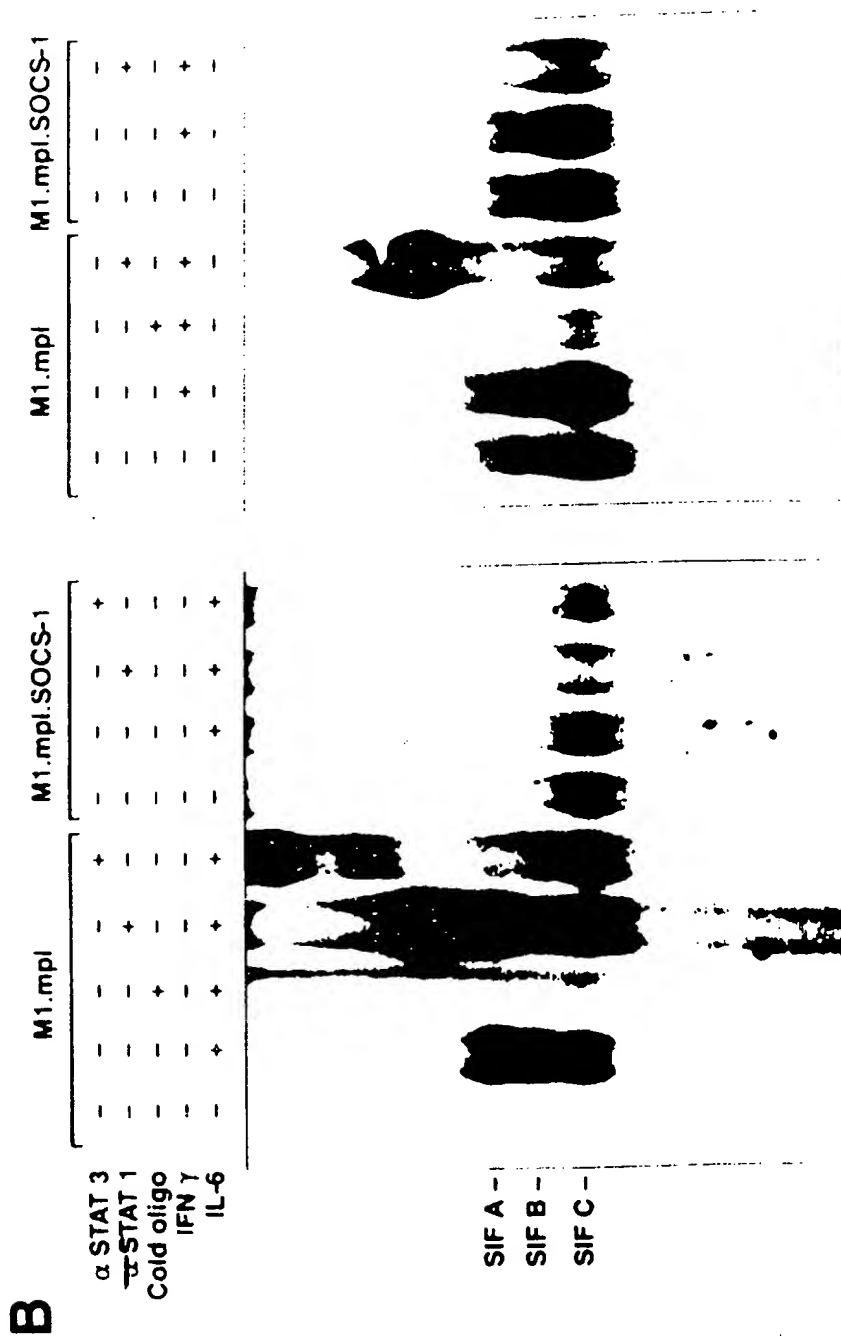


FIG 12B

20/126

FIG 13A(i)	FIG 13A(ii)
FIG 13B(i)	FIG 13B(ii)
FIG 13 C(i)	FIG 13 C (ii)
FIG 13 D	
FIG 13E(i)	FIG 13E(ii)
FIG 13 F (i)	FIG 13F(ii)

FIG 13

SUBSTITUTE SHEET (RULE 26)

21/126

A

mS	SOCS	1	WV	RRNQVAAADN	SSPAAEPRRRSE	SSSSSSSSSPA	VP
mS	SOCS	3	SH2	RTTRSKFPAAAG	SR		
mS	SOCS	2	SH2	RTTRCLEPSGNG	RDRTRSQWGTAG	LP	EEQSP
mS	SOCS	5	SH2	WVLCVQGGSCPL	LV	QI	GRRPLWAQSL
mS	SOCS	14	SH2	MDKVGKMWNNL	KYRCQNLFSH	EGGSRNEN	VE
mS	SOCS	4	SH2	SGGGPWRRAGG	GSGKSDSGLT	VEPGRGLT	ARPP
mS	SOCS	6	WD	MASFPPRVNE			
mS	SOCS	15	WD	ME			
mS	SOCS	15	WD	MGQTAL			
mS	SOCS	5	SH2	AEIPQVVEI	SI	EKDS	DSGATPGTRL
mS	SOCS	14	SH2	NFLLEKLN	TVFITL	EIVKNL	FKMAEN
mS	SOCS	5	SH2	VSSRAVGSRS	RQR	QDD	VGLCT
mS	SOCS	14	SH2	SQERQLSCSS	ELD	DDH	CCGHR
mS	SOCS	5	SH2	TFDDP	LLV	ST	DEEDR
mS	SOCS	14	SH2	IKRH	VPM	PN	DEWVSAD
mS	SOCS	5	SH2	SEEDST	TL	QL	SR
mS	SOCS	14	SH2	SEDE	II	IT	SS
mS	SOCS	5	SH2	AEIPQVVEI	SI	EKDS	DSGATPGTRL
mS	SOCS	14	SH2	NFLLEKLN	TVFITL	EIVKNL	FKMAEN
mS	SOCS	5	SH2	VSSRAVGSRS	RQR	QDD	VGLCT
mS	SOCS	14	SH2	SQERQLSCSS	ELD	DDH	CCGHR
mS	SOCS	5	SH2	TFDDP	LLV	ST	DEEDR
mS	SOCS	14	SH2	IKRH	VPM	PN	DEWVSAD
mS	SOCS	5	SH2	SEEDST	TL	QL	SR
mS	SOCS	14	SH2	SEDE	II	IT	SS

FIG. 13 A (i)

22/126

A

mS OCS. 1	SH2	RPCPAVPAPA	
mS OCS. 3	SH2	
mS OCS. 2	SH2	
mC I S	SH2	
mS OCS. 5	SH2	KEKSI SLGEA APQQESSPL RENVALQLGLSPSKTFSRRNQNC A	
mS OCS. 14	SH2	GSGRASLPRLSERRRVMAVVMAAGARTAPLELSSERSVQKVPRR	
mS OCS. 4	WD		
mS OCS. 6	WD		
mS OCS. 15	WD		
mS OCS. 5	SH2	KHSCSTKTQSSLDTEKKFGGTRSGLQRRRLRRRYGSSSHQDDMDS	
mS OCS. 14	SH2	SR SADRKDG YVWSGKLSWSKSESCSESAIGTENEIPLR	
mS OCS. 5	SH2	KHLSELMLEKCPFPAGSDLLQKWHLLKQHTAVVSSDLAFRWHF	
mS OCS. 14	SH2	PKNCSGRHSPLPSKRKI HSELMDKCCPFPVRR	
mS OCS. 5	SH2	HTFEATADVNPPLYK GPKLLPPTMTETISGDGSAAPQXNCB	
mS OCS. 14	SH2	CFSHTNGDPPCVLTANSASCTGGH TGSMMNLVTNNS	
mS OCS. 5	SH2	YTHCLVPPDILLDTGGRR	
mS OCS. 14	SH2	YTHCLVPPDILLDTGGRR	

FIG. 13 A (ii)

23 / 126

B

CONSENSUS	SH2	YFH	T	O	Ebal	T	D	I	O	I
MSOCS.1	SH2	GWY	OS	HA	HL	GS	RES	FL	SV	SV
MSOCS.3	SH2	FW	TS	AN	LL	GT	SS	LV	SV	SV
MSOCS.2	SH2	FW	MT	KN	LL	AT	SS	LV	SV	SV
MSOCS.5	SH2	FW	MT	KN	LL	AT	SS	LV	SV	SV
MSOCS.14	SH2	FW	MT	KN	LL	AT	SS	LV	SV	SV
MSOCS.9	SH2	FW	MT	KN	LL	AT	SS	LV	SV	SV
MSOCS.11	SH2	FW	MT	KN	LL	AT	SS	LV	SV	SV

FIG. 13 B (i)

[illegible]

FIG. 13 C (i)

26/126

C

CONSENSUS	K	
mmsoccs-4	YN	
mmsoccs-6	CFR	
	φWD	
	AW	QGYRI VKL
	AW	QGHCVV KL
mmsoccs-4	E H	V I D C G D I V W S L A F
mmsoccs-6		D C G Q I V W G V A F
mmsoccs-4	K I	W D V Y T G K L L L L N L V D
mmsoccs-6	K I	W E V Q T G L L L L N L S
mmsoccs-4	V W	L K D D G N M V K V L R
mmsoccs-6	I W	D L N K H G K Q I Q V L S
mmsoccs-4	L W	N M D K Y T M I R K L E
mmsoccs-6	L W	S M R S Y T L I R K L E
mmsoccs-4	V W	D P H N G D L L M E F G H L F P S P T P I F A G
mmsoccs-6	M W	D P Y T G A R L R S L H H T Q L E P T M D D S D
mmsoccs-4	F Y	R I D E D C P V Q V A P
mmsoccs-6	I W	A L E L K A P V A F A P
mmsoccs-4	F W	A
mmsoccs-6	F W	T
mmsoccs-15	G W	N
mmsoccs-15	A W	E
mmsoccs-13	G W	D I S W P L G R N R L
mmsoccs-15	G W	D I G R G K L
mmsoccs-13	M X	D
mmsoccs-15	G Y	S

FIG. 13 C (ii)

3

FIG. 13 E (i)

29/126

[illegible]

FIG. 13 E (ii)

١١

FIG. 13 F (ii)

32/126

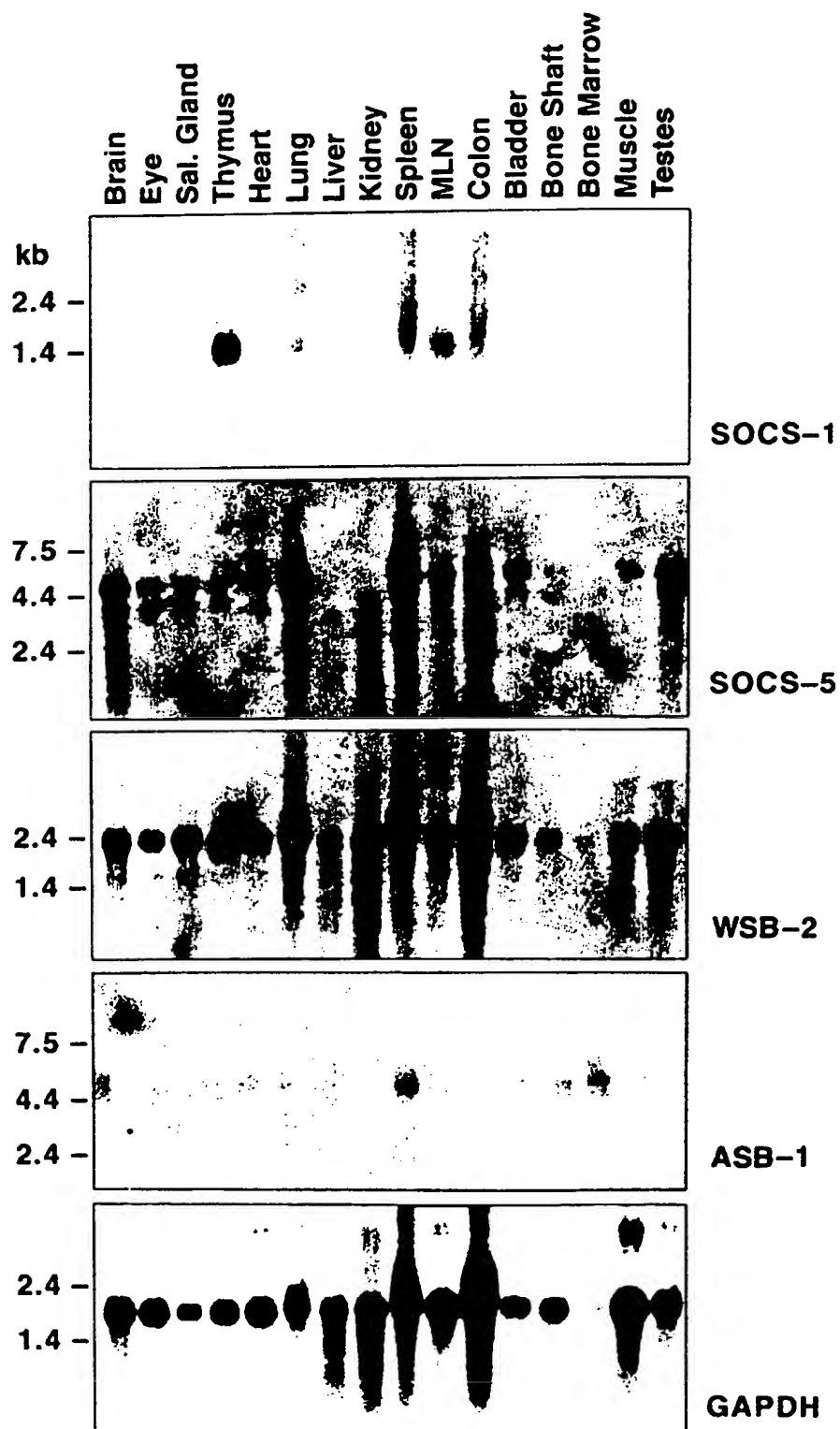
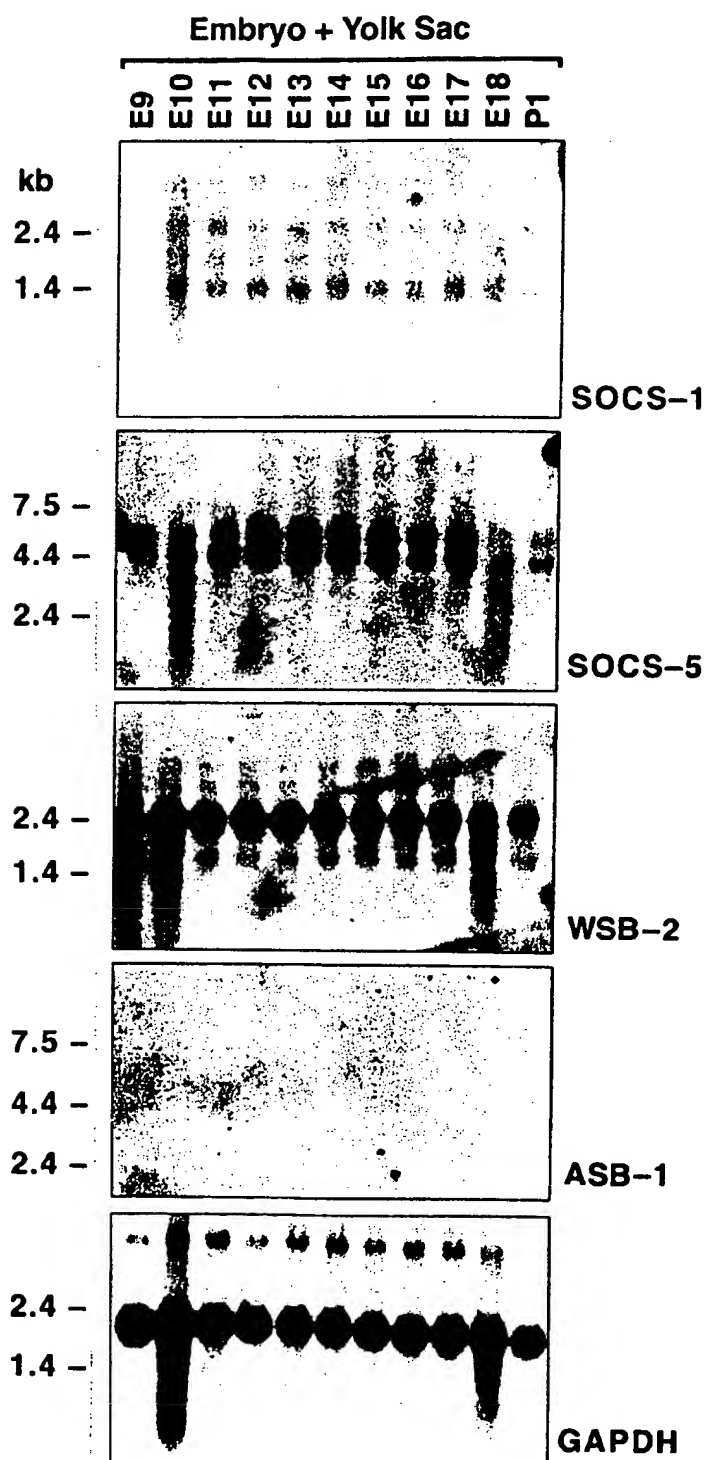
FIG 14A

FIG 14B

33/126



34/126

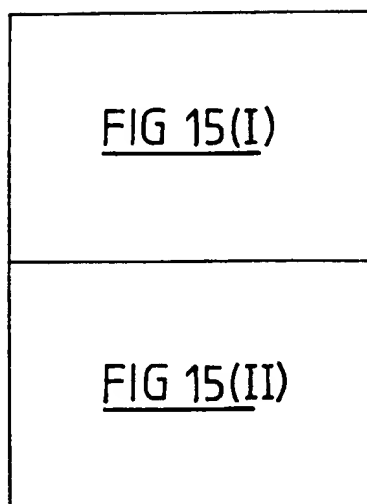


FIG 15

35/126

cgaattccggggggtgtgtgagtcctgtgagtggaaggcgccggctcttctgtct
gagtgtagcccggtggcttbtgttccaggcatccgggtgatttcttccgggcagtcgcg
agaagccgcagcgcccgccgctctctgcagtcctccacaccccgaggagacctga
gcccgcgtcacgccccctcagcccccgctgagtccttctctgtgtcgcgtccgaatc
gagttcccggaatcagacggtgcccccatagATGGCCAGCTTTCCCCCGAGGGTTAACG
AGAAAGAGATCGTGAGATCACGTACTATAGGGGAACTCTTGGCTCCAGCAGCTCCTTT
TGACAAGAAAATGTGGTGAGAACTGGACGGTTGCTTTTGCTcctGATGGTTCCCTAC
TTTGCGTGGTCACAAGGATATCGCATAGTGAAGCTTGTCCTCCGTCCAGTGCCCGTA
AGAACTTTCTTTTGCAATGGTTCCAAAAATGTTACCAATTCAAGCTGTCTAAAAATTGGC
AAGACAAAACAGTAATGGTGGTCAGAAAAACAAGCCTCCTGAGCACGTTATAGACTGT
GGAGACATAGTCTGGAGTCTTGCTTTTGGGTCTTCAGTTCAGAAAAAACAGAGTCGTT
GCGTTAATATAGAAATGGCATCGGTTCCGATTTGGACAGGATCAGCTACTCCTTGCCAC
AGGATTAAACAATGGTCGCATCAAAAATCTGGGATGTATATACAGGAAAACTCCTCCTT
AATTGGTAGACCACATTGAAATGGTTAGAGATTAACTTTTGCTCCAGATGGGAGCT
TACTCCTTGATCAGCTTCAAGAGACAAAACTCTAAGAGTGTGGGACCTGAAAGATGA
TGGAAACATGGTGAAAGTATTGCCGGGCACATCAGAAATTGGGtGtACAGTTGTGCATTc
TCTCCCgACTGTTcTATGCTGTGTTcAGTgGGCGCCAGTAAAGCAGTTTtCCTTTGGA
ATATGGATAAAcACACCATGATTAGGAAGctGGAAGGTcATCACCATGATGTTGTAGC

FIG 15(I)

36/126

TTGTGACTTTTCTCCTGATGGAGCAATTGCTAgCTACTGCATCCTATGACACTCGTGTG
TATGTCTGGGATCCACACAATGGAGACCTTCTGATGGAGTTTGGGCACCTGTCTTCCCT
CGCCCACTCCAATATTGCTGGAGGAGCAAATGACCGATGGGTGAGAGCTGTGTC'TTT
CAGTCATGATGGACTGCATGTTGCCAGCCTTGCTGATGATAAAATGGTGAgGTTCTGG
AGAAATCGATGAGGATTGTCCGGTACAAGTTGCACCTTTGAGCAATGGTCTTTTGCTGTG
CCTTTTCTACTGATGGCAGTGTTTTAGCTGCTGGGACACATGATGGAAGTGTGTATTT
TTGGGCCACTCCAAGCAAGTCCCCTAgCCCTTCAACATATATGTCGCATGTCAATCCGA
AGAGTGATGTCCACCCCAAGAAGTCCAAAACCTGCCCTGTTCCCTTCCAAAATATTGGCGT
TTCTCTCCTACCGCGGTTAGactgaagactgcctttcctggtaggcctgccagacaga
gcgccctttacaagacacacacctcaagctttacctcgtgccgaatt

FIG 15(II)

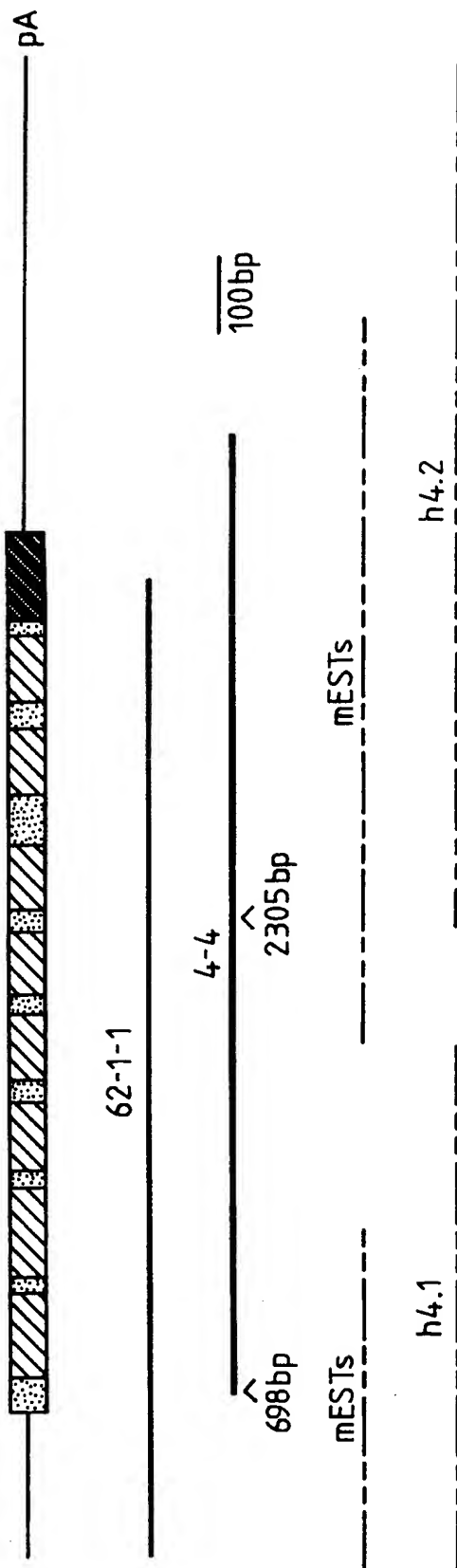
37 / 126

MASFPRVNEKEIVRSRTIGELLAPAAPFDKKCGGENWTVAFAPDGSYFAWSQGYRIV
KLVPWSQCRKNFLLHGSKNVTNSSCLKLARQNSNGGQKNKPPEHVIDCGDIVWSLAFG
SSVPEKQSRCVNI EWHRFRFGQDQLLATGLNNGRIKIWDVYTGKLLNLLVVDHIEMVR
DLTFAPDGSLLLVSASRDKTLRVWDLKDDGNMVKVLRHQNWVYSCAFSPDCSMLCSV
GASKAVFLWNMDKYTMIRKLEGGHHHDVVACDFSPDGALLATASYDTRVYVWDPHNGDL
LMEFGHLFPSPPTIFAGGANDRWVRAVSFSDHGLHVASLADDKMVRFWRIDEDCPVQV
APLSNGLCCAFSTDGSLAAGTHDGSVYFWATPRQVPSLQHCIRMSIRRRVMSTQEVOK
LPVPSKILLAFLSYRG*

FIG 16

38/126

S0CS4



SUBSTITUTE SHEET (RULE 26)

FIG 17

39/126

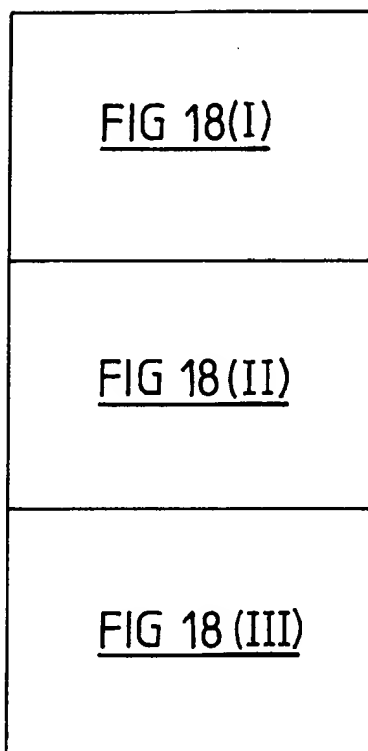


FIG 18

40/126

h4.1

CTGTCTTCCGAGCGGAGGCTGGGTACAGGGTCTATTGTCTGTGTTGACTCCG
TACTTTGGTCTGAGGCCCTTCGGGAGCTTTCCCGAGGCAGTTAGCAGAAGCCGAGCGA
CCGCCCCCGCGTCTCCTCTGTCCCTGGGCCCCGGGAGACAAACTTGGCGTCACGCC
TCAGCGTCCGCACTCTCTCTGTCTGTGTTGGTCCGCATCGTATTCCCGGAATCAGA
CGGTGCCCCATAGATGGCCAGCTTTCCCCCGAGGGTCAACGAGAAAGAGATCGTGAGA
TCACGTAATAAGTGAACTTTTAGCTCCTGCAGCTCCTTTTGACAAAGAAATGTGGTC
GTGAAAATTGGACTGTTGCTTTTGCTCCAGATGGTTCATACTTTGCTTGGTCACAAGG
ACATCGCACAGTAAAGCTTGTTCCTGGTCCAGTGCCCTTCAGAACTTTCTCTTGCA
GGCACCAGAAATGTTACCAATTCAAGCAGTTTAAGATTGCCAAGACAAAATAGTGATG
GTGGTCAGAAAAATAAGCCCTCGTGACATATTATAGACTGTGGAGATATAGTCTGGAGT
CTTGCTTTTGGGTCAATCAGTTCAGAAAAACAGAGTCGCTGTGTAATATAGAATGGC
ATCGCTTCAGATTTGGACAAGATCAGCTACTTCTTGCTACAGGGTTGAACAATGGGCG
TATCAAAATATGGGATGTATATMCAGGAAACTCCTCCTTAACTTGGTAGATCATACTG
AAGTGGTCAGAGATTAACTTTTGCTCCAG

FIG 18(I)

41/126

h4.2

CTCTGTATGTCTGAATGAAGCTATAACATTTGCCCTTTTATTTGCAGGTTTTCCTTTGG
AATATGGATAAATACACCATGATACGGAACTAGAAGGACATCACCATGATGTGGTAG
CTTGTGACTTTTCTCCTGATGGAGCATTACTGGCTACTGCATCTTATGATACTCGAGT
ATATATCTGGGATCCACATAAATGGAGACATTTCTGATGGAATTTGGGCACCTGTTTCCC
CCACCTACTCCAATATTTGCTGGAGAGCAAAATGACCGTGGTACGATCTGTATCTT
TTAGCCATGATGGACTGCATGTTGCAAGCCTTGCTGATGATAAAATGGTGAGGTTCTG
GAGAAATTGATGAGGATTATCCAGTGCAAGTTGCACCTTTTGAGCAATGGTCTTTTGCTGT
GCCCTCTCTACTGATGGCAGTGTTTTAGCTGCTGGGACACATGACGGAAGTGTGTATT
TTTGGGCCACTCCACGGCAGGTCCCTAGCCCTGCAACATTTATGTCCGCAATGTCAAATCCG
AAGAGTGATGCCCCACCCCAAGAAAGTTCAGGAGCTGCCGATTCCCTTCCAAGCTTTTGGAG
TTTCTCTCGTATCGTATTTAGAAGATTCTGCCCTTCCCTAGTAGGGACTGACAGAA
TACACTTAACACAACCTCAAGCTTTTACTGACTTCAATTATCTGTTTAAAGACGTA
GAAGATTATTAATTTGATATGTTTCTTGTAAGTTCATTTTGATCAGTTGAGCTTTTAA
AATATATTATAGACAAATAGAAAGTATTTCTGAACATATCAAAATATAAATTTTAA
AGATCTAACTGTGAAAACATACATACCTGTACATATTTAGATATAAAGCTGCTATATGT
TGAATGGACCCCTTTTGCTTTTCTGATTTTCTAGTTCTGACATGATATATTTGCTTCAGT
AGAGCCACAAATATGTATCTTTTGCTGTAAAGTGCAAGGAAATTTTAAATTTCTGGGACAC

FIG 18(II)

42/126

TGAGTTAGATGGTAAATACTGACTTACGAAAGTTGAATTGGGTGAGCGGGGCAAATCA
CCTGAGGTCAGCAGTTTGAGACTAGCCTGGCAAACATGATGAAACCCCTGTCTCTACTA
AAAATACAAAAAATAAAAAA

FIG 18(III)

S0CS5

43/126

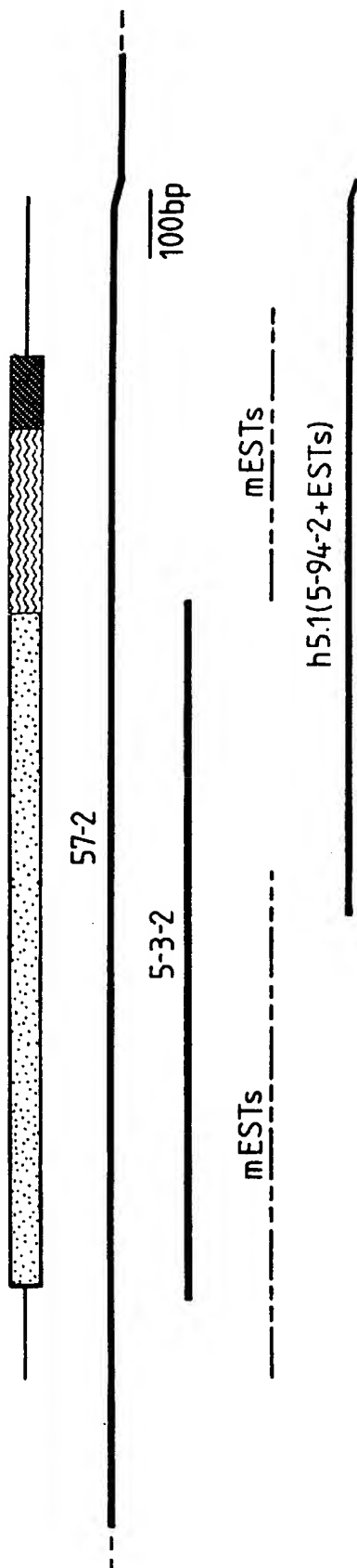


FIG 19

44/126

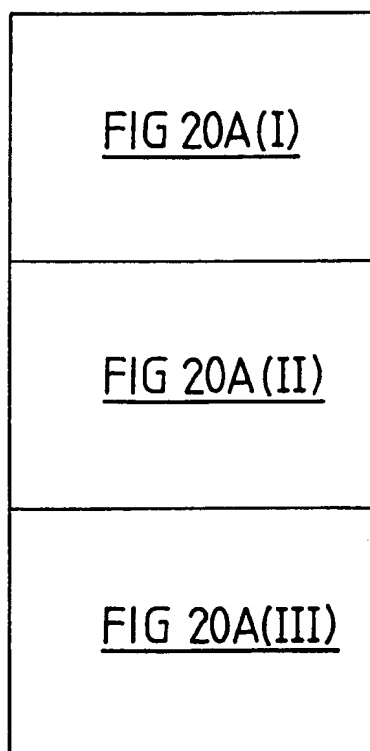


FIG 20A

45/126

cggcacgagccgggctccgtccggaggaaagcgaggctgcgccgccggcccgccagagagc
ggaggacgggagmcgcggcggtcgcgctcgccctgcgctgactgcgctgccccggcc
catccttgccctggccgcaggtgccctggatgagggccgcgcgctgtcccgccgctga
gtgtcccccggtcgcgccggcctgcctcaagcgccgctctccttgccccgggtc
cccgtttcccccggcagtcctcctccggtggcgccctccgcacctcggcgccagggc
gcacggccctcgggccgggatggatccgcgggaagaggaagacaagccgggcgttga
gcccctgcgcacggtgcgcgcgcgtagtgggagcttactcgcagtaggctctcgctc
ttctaatacaATGGATAAAGTGGGAAAATGTGGAACAACCTTAAAAATACAGATGCCAGAA
TCTCTTCAGCCACGAGGAGGAAGCCGTAATGAGAACGTGGAGATGAACCCCAACAGAT
GTCCCGTCTCAAGAGAAAAGCATCAGTCTGGGAGAGGCAGCTCCCCAGCAAGAGAG
CAGTCCCCTTAAGAGAAAATGTTGCCCTTACAGCTGGGACTGAGCCCTTCCAAGACCTTT
TCCAGGCGGAACCAAACTGTGCCGCAGAGATCCCTCAAGTGGTTGAAATCAGCATCG
AGAAAGACAGTGACTCgGGTGCCACCCAGGAACGAGGCTTGACGGAGAGACTCCTA
CTCGCGCACGCCCGTGGGGAGGAAAGAAACATTCTGTTCACAAAGACCCAG
AGTTCAATTGGATACCGAGAAAAGTTTGGTAGAACTCGAAGCGGCCTTCAGAGGCGGAG
AGCGCGCTATGGAGTCAGCTCCATGCAGGACATGGACAGCGTTTCTAGCCGCGCGGT
CGGGAGCCGCTCCCTGAGGCAGAGGCTCCAGGACACGGTGGGTTTGTGTTTCCCCATG
AGAACTTACAGCAAGCAGTCAAAGCCACTCTTTTCCAATAAAAGAAAAATACATCTTT

FIG 20A(I)

46/126

CTGAATTAAATGCTGGAGAAATGCCCTTTTCCCTGCTGGCTCGGATTTAGCACAAAAGTG
GCATTTGATTAAACAGCATACCGCCCCCTGTGAGCCCCACACTCAACATTTTGTGATACA
TTTGATCCATCACTGGTGTCTACAGAAGATGAAGAAGATAGGCTTCGCGAGAGAAGAC
GGCTTAGTATCGAAGAAGGGGTGGATCCCTCCCAACGCACAATAACACACCTTTTGA
AGCTACTGCACAGGTCAACCCATTGTATAAGCTGGGACCAAAAGTTAGCTCCTGGGATG
ACAGAGATAAGTGGAGATGGTTCTGCAATTCCACAAGCSAATTGTGACTCAGAAAGAGG
ATTCAACCACCTATGTCTGCAGTCACGGAGGCAGAAAGCAGGCCAGGTGTCCGGGGA
CAGCCACGCGCACGTTAGCAGACAGGAGCTTGGAAGTTCAACGCAGATCGATATAC
ATACACTGCCCTCGTGCCAGATTTTGCTTCAGATCACAGGGAATCCCTGTTACTGGGCG
TGATGGACCGATACGAGGCCGAAGCCCTTCTAGAAGGGAACCGGAAGCACGTTCTT
GCTCAGGACTCTGCACAGGAGACTACCTCTCTCTGTGTGAGCTTCGCGCGCTACAAC
AGGTCCTCTGCACGCCGGATCGAGCAGTGGAAACCACAACCTTCAGCTTCGATGCCCATG
ACCCCTGCGTGTTCACCTCCTCCACwGTCAACGGGGCTTCTCGAAACACTATAAAGACCC
CAGCTCTTGCAATGTTTTTTGAACCGTTTGCTAACGATATCACTGAATAGAACTTTCCCT
TTCAGCCTGCAGTATATCTGCCGCGCAGTGATCTGCAGATGCACCTACGTATGATGGGA
TTGACGGGCTCCCGCTACCGTCGATGTTACAGGATTTTTTAAAGAGTATCATTTATAA
ACAAAAAGTTAGGGTTCGCTGGTTAGAACGAGArCCAGTCAAAGCAAAGTAActctg
tccccaaaggcactaaagtctgctcctcccgtgcatcmgaactgcaccccatagg

FIG 20A(II)

47/126

raggcagtcagctgctaggatttcccacccagaatgggagcttagtcattagcctctg
ccctatggggtccgctgttcttcagacaaaggctgcctagggacagcaagatggcttgc
aggtgtcgggtggctgtgacaactgaggaggcaactctggggcatcttgctatgaag
aattctatttcttaccgaagaacaaattattaatatggatgggtatttcaatagtgt
gactaatgtttgaaattatttttcttaagaatttttctataaccttcagaaaaagtag
tgatgtttgtagttactataaatcaagctttgaaagttcaaaacaacaagttaaata
aaagactaccttccttttagagaaaaacaaatgcaagtttcccagccacaggcattgt
gcactgttaatgttngcttgttatcagctccttttctctcc

FIG 20A(III)

48/126

MDKVGKMWNNLKYRCQNLFSHEGGSRNENVMNPNRCPVKEKSI SLGEAAPQQESSP
LRENVALQLGLSPSKTFSRRNQNCAAEI PQVVEISIEKSDSGATPGTRLARRDSYSR
HAPWGGKKKHSCSTKTQSSLDTEKKFGRTRSGLQRRRERYGVSSMQMDSVSSRAVGS
RSLRQRLQD TVGLCFPMRTYSKQSKPLFSNKRKIHLSELMLEKCPFPAGSDLAQKWHL
IKQHTAPVSPHSTFFDTPSLVSTEDRRLRERRRLSIEEGVDPPPPNAQIHTFEAT
AQVNPLYKLGPKLAPGMTEISGDGSAI PQXNCDSEEDSTTLCQSRQKQRQVSGDSH
AHVSRQGAWKVHTQIDYIHCLVPDLLQITGNPCYWGVMDRYEAALLEGKPEGTFLLR
DSAQEDYLFVSFRRYRNSLHARIEQWNHNFSDAHDPCVFHSSXVTGLLEHYKDPSS
CMFFEPLLTISLNRTPFESLQYICRAVICRCTTYDGDGLPLPSMLQDFLKEHYHKQK
VRVRWLERXPVKAK*

FIG 20B

49/126

GATTAAACAGCATACAGCTCCTGTGAGCCCAACATTCAACATTTTTTGTATAC-TTTGATCCATCTTTGGTTT
CTACAGAAAGATGAAGAAGATAGGCTTAGAGAGAGAAAGGCGCTTAGTATTGAAGAAGGGTTGATCCC
CCTCCCAATGCACAAATACATAACATTTGAAGCTACTGCACAGGTTAATCCATTATwTAAACTGGGACCA
AAATTAGCTCCTGGAATGACTGAAATAAGTGGGGACAGTTCTGCAATTCCCAAGCTAATTGTGACTCG
GAAGAGGATACAACCAACCCCTGTGyTTGCAGTCAAGGAGGCAGAAAGCAGCGTCAGATATCTGGAGACAGC
CATACCCATGTTAGCAGACAGGGAGCTTGGAAAGTCCACACACAGATTGATTACATAACACTGCTTCGTG
CCTGATTTGCTTCAAATTACAGGGAATCCCTGTTACTGGGGAGTGATGGACCCGTTATGAAGCAGAAAGCC
CTTCTCGAAGGGAACCTGAAGGCACGTTTTTTGCTCAGGGAAGCTCTGCGCAAGAGGACTACTTCTTCTCT
GTGAGCTTCCGCCGATACAACAGATCCCTGCAATGCCCGAATTGAGCAGTGGAAATCACAACCTTTAGTTTC
GACGCCCATGACCCCGTGTGTAATTTCACTCCTCCACTGTAAACGGGACTTTTAGAACATTATAAAGATCCCCA
GTTTCGTGCATGTTTTTTTGAAACCAATTGCTTACTATATCACTAAATAGGACTTTCCCTTTTAGCCTGCAGTAT
ATCTgTcGCGCGGTAACTCTGCAGGTGCACTACGTATGATGGAATTGATGGCTCCCTCTACCCCTCAATGT
TACAGGATTTTTTAAAAGAGTATCATTTATAAACAAAAAGTTAGAGTTCGCTGGTTGGaACGAGAACCCAG
TCAAGGCAAGTAAACTCTCCGGTCCCCAAAGGgTGTTAACTAGGTCCGCTTTCATGTGCAATCAGACAGT
ACACCTATAGCAAGCACACGTAGCAGTGTTAGGCTTTTTCATACAGTATGTAAAGcTTAGTGTAGTATCT
GTCAGaIGCTACCTGCTGTACTTATTCAGATAAACATGGhGCCCTATTGGAAACAATAGcGGATAGAGCTAC
AGGTGTTTCAGTAAGACTACAAAAACATTTTGCCTATTTCGCTAACAGTTTGGTTTTTAATGGCTGTGGtA
TTTTAGTGAGGCAACTCTGGGGCAATTTGTtATGAAGAAATG

FIG 21

50/126

SOCS6

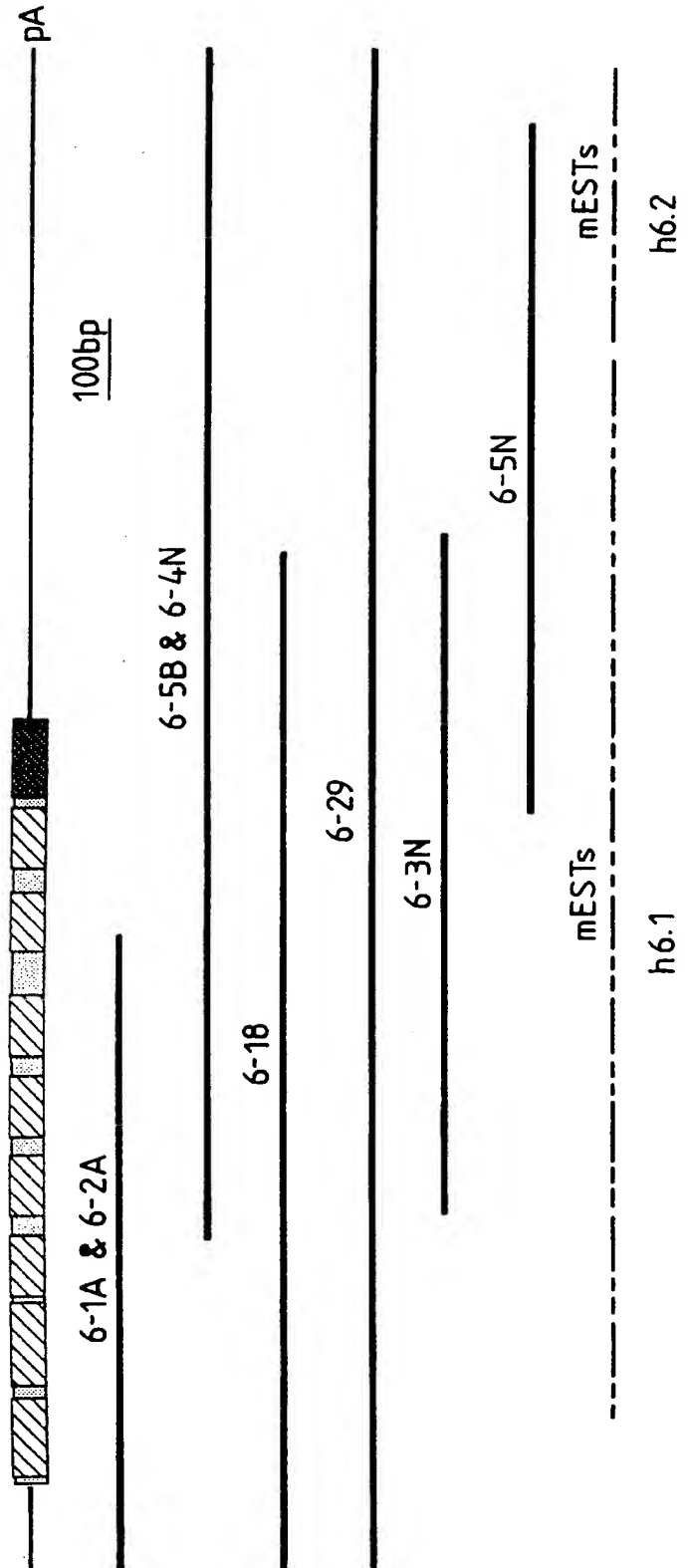


FIG 22

51/126

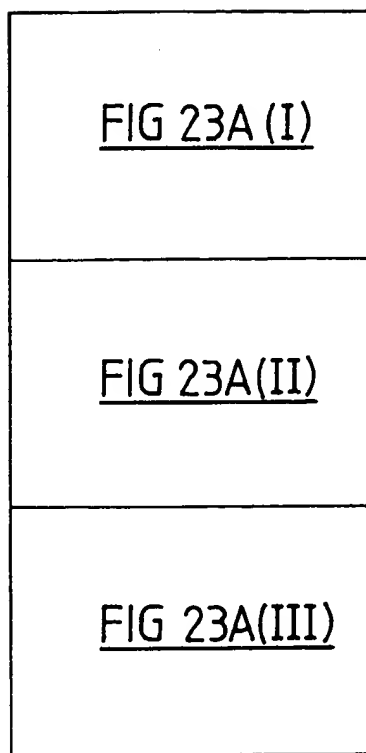


FIG 23A

52/126

ggcacgagcggtggtggcggcgggcgggcgggcgggcgggcgggcggaATGA
AGGCCACGGCCCTGGGGCTGAGGCGCCCGCCGCTGGGGCGGGCGCGTCTCA
TGGAGGCCGGAGAGCGCGCTGCTGGCTGAACCTCAAGCCTGGCGCCCCCACC
GTTCGACTGGAAGTCAAGCTGCGAGACCTGGAGCGTGGCCTTCTCGCCAGACGGTTCC
TGGTTCGCCCTGGTCTCAAGGACACTGCGTGGTCAAGCTGGTCCCCCTTAGAGG
AACAGTTCATCCCTAAAGGATTCGAAGCCAAGAGCCGAAGCAGCAAGAATGACCCAAA
AGACGGGGCAGTCTGAAGGAGAAAGACGCTGGACTGTGGCCAGATTGTGTGGGGCTG
GCCTTCAGCCCGTGGCCCTCTCCACCCAGCAGGAACCTCTGGGCACGTCAACCATCCCC
AGCGCCTGATGTTCTTGCCCTGATCCTGGCCACAGGTCTCAACGATGGGCAGATCAA
GATTGGGAGGTACAGACAGGCCCTCCTGCTTCTGAATCTTCTGGCCACCAAGACGTC
GTGAGAGATCTGAGCTTCACGCCCCAGCGGCAGTTTGATTTTGGTCTCTGCATCCCCGGG
ATAAGACACTTCGAATTTGGGACCCTGAATAAaCACGGTAAGCAGATCCAGGTGTTATC
CGGCCATCTGCAGTGGGTTTACTGCTGCTCCATCTCCCCCTGACTGTAGCATGCTGTGC
TCTGCAGCTGGGGAGAAAGTCGGTCTTTCTGTGGAGCATGCGGTCTACACACTAATCC
GGAAACTAGAAGGCCACCAAGCAGTGTGTCTCCTGTGATTTCTCTCCTGATTCAGC
CTTGCTTGTCACAGCTTCGTATGACACCAAGTGTGATTAATGTGGACCCCTACACCGGC
GcGAGGCTGAGGTCACCTCATCACACACAACCTTGAAACCACCATGGATGACAGTGACG
TCCACATGAGCTCCCCTGAGGTCCGTGTGCTTCTCACCTGAAGGCTTGTATCTCGCTAC

FIG 23A(I)

53/126

GGTGGCAGATGACAGGCTGCTCAGGATCTGGGCTCTGGAACCTGAAGGCTCCGGTTGCC
TTTGCTCCGATGACCAATGGTCTTTGCTGCACGTTCTTCCACACAGGTGGAATTATTG
CCACAGGACGAGAGATGGCCATGTCCAGTTCTGGACAGCTCCCCGGGTCTGTCCCTC
ACTGAAGCACTTATGCAGGAAAGCCCTCCGAAGTTTCCTGACAAACGTATCAAGTCCTA
GCACTGCCAATCCCCAAGAAGATGAAGAGTTCCCTCACATACAGGACTTCTTAGCagt
gccggctccccacctcctgcagcagcagcagtaacaaggactggctaggatggagtc
aggcagctcacactggaccagtggtggaccttccttccctcccatggcatgtgcaagtag
gtctgctgaccccaactctgtggtgccggccttacctcgtcttcacccgtggtgagc
agccttcgtcagtcctagtgtgtgtgaaagccaaagtcagttgtggatgttgctggggtgta
ataaaggcaagcgggtccagagacctctctggtggcgccaaagccacactcccttaac
tgggaagtacctgccacgtagggcatttctgctgccctattccagccagcggctgcat
ggtttgaagtccctccgtgtgtggtcagaagaactctggtgttggttccctgctcagg
tgcgctggactgggctgagctcctcacatacactagtgccggctttgtttcctgt
aaacagtgggtgcattgtgtagagaagtaacaagcaggtattcagatcatacagaggagg
cgttccctcgggtgcattgacgggtcagatggccatttatcagcataatttatgtattttc
tcagcacatagtaaggtaacaactgtgttttctcaattgtctcgaaaaaacagagttct
taagtggcccagttgtggagccaaagtcctaaagtcgtgtggagtcagtgctgacatcact
ggcttgctgtctgtcacatgtgtttgtctctgtgcttgacctcatgggatgtacc

FIG 23A(II)

54/126

ctccagttcaactgccccaaaacagacagcccccttccaagcacccgttctttgacagcgg
tagcagctacctattcaagacgcctcacacaaaaatctgccttagaaaagttaatatatt
ttaaatattttaaaagaactcaacatcttattctttggcctttctttaattgatgct
ttatggaggcagtgtaacattgtacagtgtagcatagaggagtcctcctctatttga
agaacaatgcaaaatgaggctttcattgaagggaataaaaaaa

FIG 23A(III)

55/126

MEAGEEPLLLAELKPGRPHQFDWKSSCETWSVAFSPDGSWFAWSQGHCVVKLVPPWPLE
EQFIPKGFEEKSRSSKNDPKGRGSLKEKTLDCGQIVWGLAFSPWPSPPSRKLWARHHP
QAPDVSCILLATGLNDGQIKIWEVQTGLLLNLNSGHQDVVRDLSFTPSGSLILVSASR
DKTLRIWDLNKHGKQIQVLSGHLQWVYCCSISPDSCMLCSAAGEKSVFLWSMRSYTLI
RKLEGHQSSVSCDFSPDSALLVTASYDTSVIMWDPYTGARLRSLHHTQLEPTMDDSD
VHMSSLRVCFSP EGLYLATVADDRLLRIWALELKAPVAFAPMTNGLCCTFFPHGGII
ATGTRDGHVQFWTAPRVLSSLKHLCKRKALRSFLT~~TTYQVLA~~LPIPKMKKEFLT~~YRTF~~*

FIG 23B

56/126

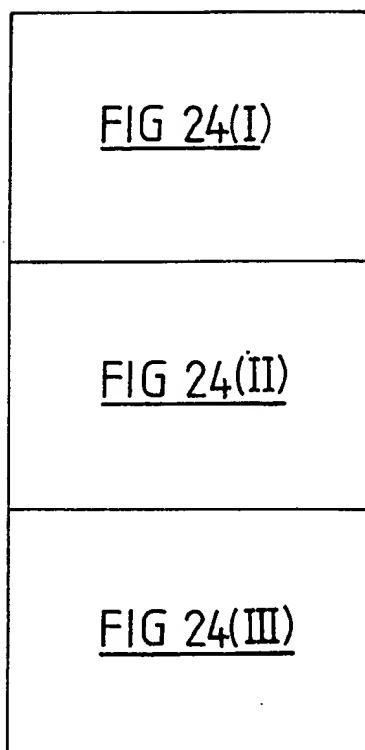


FIG 24

57/126

h6.1

GACACTGCATCGTCAAACTGATCCCCCTGGCCGTTGGAGGAGCAGTTTCATCCCTAAAGG
GTTTGAAGCCAAAGCCGAAGTAGCAAAAATGAGACGAAAGGGCCGGGAGCCCAAAA
GAGAAGACGCTGGACTGTGTCAGATTGCTCTGGGGCTGGCCCTTCAGCCCTGTGNCCTT
CCCCACCCAGCAGGAAGCTCTGGGCACGCCACCACCCCAAGTGCCCCGATGTCTCTTG
CCTGGTTCTTGCTACGGGACTCAACGATGGGCAGATCAAGATCTGGGAGGTGCAGACA
GGGCTCCTGCTTTTGAACTTTCCGGCCACCAGATGTCGTGAGAGATCTGAGCCTTCA
CACCAGTGGCAGTTTGATTTTGGTCTCCGGCTCACGGGATAAGACTCTTCGCATCTG
GGACCTGAATAAACACGGTAACACAGATTCAAGTGTTATCGGGCCACCCTGCAGTGGGT
TACTGCTGTTCCATCTCCCCAGACTGCAGCATGCTGTGCTCTGCAGCTGGAGAGAAAGT
CGGTCTTTCTATGGAGCATGAGGTCCCTACACGTTAATTCGGAAGCTAGAGGGCCATCA
AAGCAGTGTGTCCTTGTGACTTCTCCCCGACTCTGCCCTGCTTGTACGGCTTCT
TACGATACCAATGTGATTATGTGGGACCCCTACACCGCGGAAAGGCTGAGTCACTCC
ACCACACCCAGGTTGACCCCGCCATGGATGACAGTGACGTCCACATTAGCTCACTGAG
ATCTGTGTGCTTCTCTCAGAAAGGCTTGTAACCTTGCCACGGTGGCAGATGACAGACTC
CTCAGGATCTGGCCCTGGAACTGAAACTCCCATTTGCAATTGCTCCTATGACCAATG
GGCTTTGCTGGCACATTTTTTCCACATGGTGGAGTCATTGCCACAGGGACAAGAGATG
GCCACGTCCAGTTCTGGACAGCTCCTAGGGTCTCTCCTCACTGAAGCACTTATGCCG

FIG 24(I)

58/126

GAAAGCCCTTCGAAGTTTCCCTAACAACTTACCCAAGTCCCTAGCACTGCCAATCCCCAAG
AAAATGAAAGAGTTCCCTCACATACAGGACTTTTAAAGCAACACCCACATCTTGTGCTTC
TTTGTAGCAGGGTAAATCGTCCTGTCAAAGGGAGTTGCTGGAATAATGGGCCCAAACAT
CTGGTCTTGCAATTGAAATAGCAATTTCTTTGGGATTGTGAATAGAAATGTAGCAAAACCA
GATTCAGTGTACTAGTCATGGATTTTTC

FIG 24(II)

59/126

h6.2

ACCATGGTTCCAAGWTCCTCTCCYKCCCTGTGGTCMRAAGTTGCYCCGAAATGTTGGGC
CCAAGTGCCCTTTTCYCTCCTTGGGCCCTCCCTTCTGACCTGCAGGACAGTTTTCYGG
AGCCCAATTTGGTATGAGGTATTAAWTTAGCCCTTAACATAAATTACAGGGGACTCAGAGG
CCGTGCTCCTGACCGATCCAGACACTATTTTTTTTTTTTTTTTAAACAATGGTGTC
ATGTGCAGGAAATGACAAATTTGTATGTCAGATTATACAAGGATGTATTCTTAAACCG
CATGACTATTCAGATGGCTACTGAGTTATCAGTGGCCATTATTAGCATCATATTAT
TTGTATTTTCTCAACAGATGTTAAGGTACAACACTGTGTTTCTCGATTATCTAAAAAC
CATAGTACTTAAATTGAAAAA

FIG 24(III)

60/126

SOCS-7

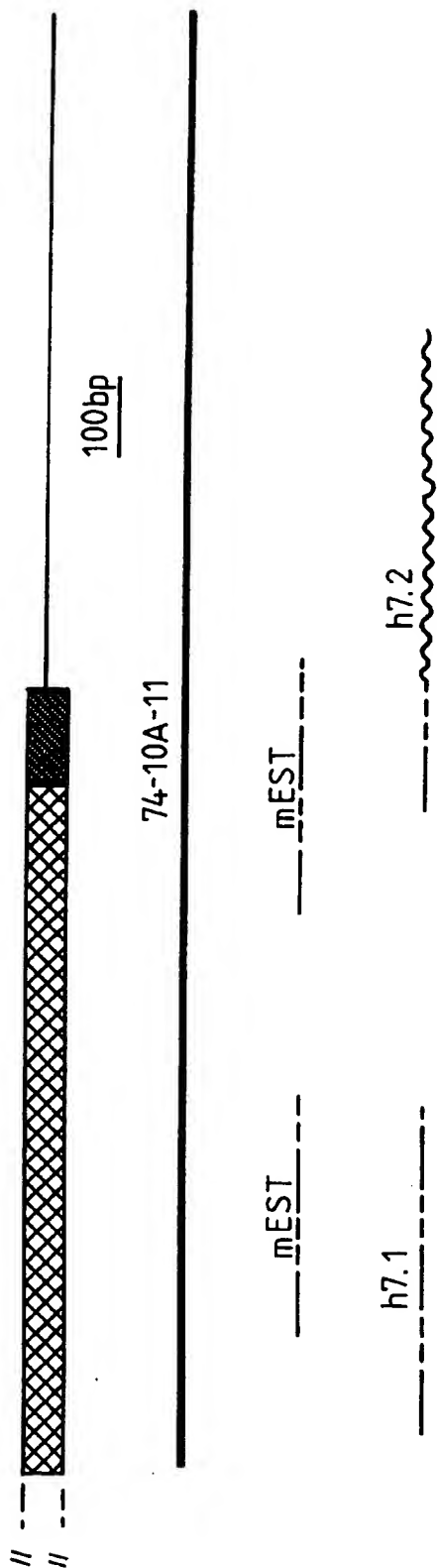


FIG 25

61/126

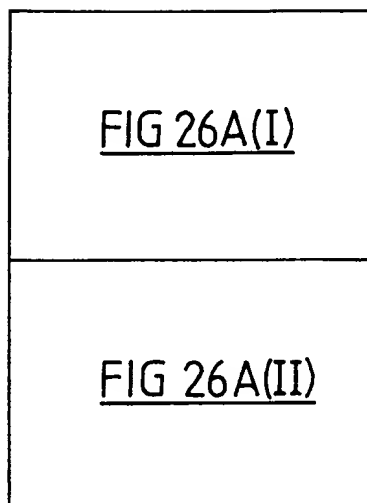


FIG 26A

62/126

GGCACGAGCGGGGTCAGGGCGGAGGCTGAGGACCAAGTAGGCATGGCGAGGGCGGG
ACCGCCCCGATGGACGGGCGGCCCCGGGACCCGCAGGTCCTAATCTGAAGGAGTGGC
TGAGGGAGCAGTTCTGTGACCATCCACTGGAGCACTGTGACGATACAAGACTCCATGA
TGCAGCCTATGTAGGGACCTCCAGACCCCTCAGGAACCTACTGCAAGAGGAGAGCTAC
CGGAGCCGCATCAATGAGAAAGTCTGTCTGGTGCTGGGGCTTGCTCCCTGCACACCAC
TGAGGATCGCAGCCACTGCAGGCCCATGGGAACCTGTGTGGACTTCCCTCATACGCAAAGG
GGCCGAGGTGGACCTGGTGGATGTCAAGGGGCAGACTGCCCTGTATGTGGCTGTAGTG
AACGGGCACCTTGGAGAGCACTGAGATCCTTTTGGAAAGCTGGTGCTGATCCCAACGGCA
GCCGGCACCCAGCAGCACTCCTGTGTACCATGCCCTYTCGTGTGGGTAGGGACGACAT
CCTGAAGGCTCTTATCAGGTATGGGGCAGATGTTGATGTCAACCATCATCTGAATTCT
GACACCCGGCCCCCTTTTTCACGGCGGCTAACCTCCTTGGTGCTGTCTCTATACA
TCAGTGCTGCCCTACCATAAACCTTCAGTGCTTCAGGCTGCTCTTGCCAGGCTGGGGCAA
TCCTGACTTCAATTGCAATTGGCCCCGTGTCAACACCCAGGAGTTCACAGGGGATCCCCCT
GGGTGTGCATGGATGCTGTCCCTGCGCCATGGCTGTGAAGCAGCCCTTCGTGAGTCTGT
TGGTAGAGTTTGGAGCCAACCTGAACCTGGTGAAGTGGGAATCCCTGGGCCCAGAGGC
AAGAGGCAGAAAGATGGATCCTGAGGCCCTTGCAGGTCTTTAAAGAGGCCCAGAAAGT
ATTCCCAGGACCTTGCTGAGTTTGTGCCGGGTGGCTGTGAGAAGAGCTCTTGGCAAAT
ACCGACTGCATCTGGTTCCCTCGCTGCCGTGCCAGACCCCATAAAGAAGTTTGTGCT

FIG 26A(I)

63/126

TTATGAGTAGcattcacatgcagtgcctgactgcaatgtggaagccgatacacctgcagt
gaaaactgacacagactctggcatcctgggaaccatggcctgtgctgccagcttgatc
cttggctgtcagtgaagaaaaaacggctgtgttctcttggactgtgattctatctcag
gtgcttgggcatcgaaacgctccttgagtcattgtcaactgagaggcacatacaaaact
taattttgttcctcttcagtcctctctgttttggaattcttcctggcaatgtgtgcagca
tgggctgagcctgggtgattgccctagtggggaaggctttttctccaggctatgcatac
tatttatgttcctactttgcaatttatgttctttaaggcttgatatcaaaacagaa
agaggtttgttaagaaaagatataggagaaaaggaaattccggttccgtgcacttgcta
gcctgtttccttgccctgggtttgtctgtctatgtcctgctggcgacatcccttctct
ttgctgccactgttctattttgggagttgtcttccgtctaaagatggcttctgggggttc
tatcttatcgacagaggtcccagaaacagtggttcataggggcaccatctgctctgccaa
gggttttctgatgtcttaccctggggatcttcagacagtggttacctttaggagaccc
acctggaaactaaccattaaagtgactgcccacattcagatcaggggaccatcttaatagt
actcactgccagtcctcacaaagagaagatgacacgggtgctctcttcagacactccca
tacagggaagttggaaaaatgtcttggtcacctgggtgttcccaggctacaacttcttg
gtgtccactaaraccagratactcctagtttttgggttgactgttccctccccactt
tccttgaancccaatgcccntttgtktnggttgcttccctaaaaakt

FIG 26A(II)

64/126

....ARGGVRAEAEQVGM AEGGTGPDGRAGPGPAGPNLKEWLREQFCDHPLEHCDDT
RLHDAAYVGD LQTLRNLLQEE SYRSRINEKSVWCCGWL PCTPLRIAATAGHGNCVDFL
IRKGA EVDLV DVKGTALYVAVVNGHLESTEIL LEAGADPN GSRHHRSTPVYHAXRVG
RDDILKALIRY GADV DVNHHNLNSDTRPPFSRRLTSLVVCPLYISAA YHNLQCFRLLLQ
AGANPD FNCN GPVNTQEFYRGSPGCVMDAVLRHGCEAA FVSLLVEFGANLNLVKWESL
GPEARGRRKMDPEALQVFKEARSI PRTL LSLCRVAVRRALGKYRLHLVPSLPLPDPIK

KELLYE*

FIG 26B

65/126

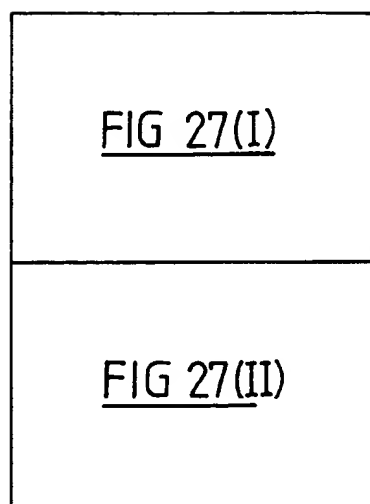


FIG 27

66/126

h7.1

GCATCCATGGCGGAGGGCGGCAGCACGACGGGCGGGCAGGGCCGGGCTCCGCAGGTCCG
TAATCTGAAGGAGTGGCTGAGGGAGCAAATTTTGTGATCATCCGCTGGAGCACTGTGAG
GACACGAGGCTCCATGATGCAGCTTACGTCGGGGACCTCCAGACCCTCAGGAGCCCTAT
TGCAAGAGGAGAGCTACCGGAGCCGCATCAACGAGAAAGTCTGTCTGGTGTGTGGCTG
GCTCCCCTGACACACCGTTGCCAATCGCGGCCACTGCAGGCCATGGGAGCTGTGTGGAC
TTCCCTCATCCGGAAGGGGCCGAGGTGGATCTGGTGACGTAAGGACAGACGGCCC
TGTATGTGGCTGTGTAACGGGCACCTAGAGAGTACCCAGATCCTTCTCAGAGCTGG
CGCGGACCCCAAC

FIG 27(I)

67/126

h7.2

GAGGAAGAAAAAGTGGACCCCTGAGGCCCTTGCAGGTCTTTAAAGAGGCCAGAGTGT
TCCCAGAACCTTGCTGTGTCTGTGCCGTGTGGCTGTGAGAAAGAGCTCTTGGCAAAMAC
CGGCTTCATCTGATTCCCTTCGCTGCCCTCTGCCAGACCCCATAAAGAAAGTTTCTACTCC
ATGAGTAGACTCCAAGTGTGCTGCCGTTGATTCCAGTGAGGGAGAAAGTGATCTGCAGGG
AGGTGGACACCGAGCCCTGAGTGTCTGTGCTGCTGGTCTCCTGATGGCTGTTGCTG
CAGAAATGTCCTCGTAGACTGTCAATTGCTCCTCAGGTGCCCTGGGCCGCTGAACAGTC
CTTGGGTCAATTGTCACTGAGAGGCTTATACATAAAGTTATTATGTTTTCCTCCAAAGTT
CTCTGTTCTGGATTTCAGTTGCATATTAAATGTAAACGGGCCATGGGGTATGTACATGT
AGGGGCTGAGGTTGAGGCCCTACTAATTTCCTGTAGGGAAGACTCCACGACTTCTTGG
AACTGTGCTTCTCTTTATTTTCTACTTCTCAATTGTGATGGTTCGATTAAAGCCTTCT
AGTATCTCAATGAAAA

FIG 27(II)

68/126

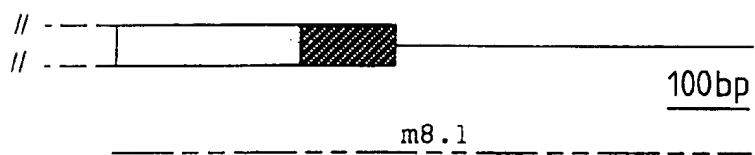


FIG 28

69/126

CTGATGTCCGCAATTCTGAAGTTGGACACCACTGCTGGCTGCCCTGTGACATCCGCTG
TCAATCCCCAAAGGATGCTGAGGCCACCACCAACCGCTGTTTCAACTGTGCCGCTTG
CTGCTGTCTGTGGGGCAGATGCTGATGAATACATACCGTGTAATTCAGCTTCCCTGAG
GAGGCCAAGGgCTTGTTGCCACCAGAGATTCTACAGAAAGTACCATGGATTCTACTCTT
CCCTCTTTGGCTTGGTGAGGCAGCCCAGGTGCTGCAGCATCTCTGCCGTTGTGCGCT
CCGCAGTCACCTGGAGGGCTGTCTGCCCCATGCACCTACCGCGCCCTTCCCCTGCCACCG
CGCATGCTCCGCTTCTGCAGCTGGACTTTGAGGATCTGCTCTACTAGgcttgctgcc
ctgtgaacaaagcagacccccaccaccccccaaggcatctctcagcaatgaatgatg
caaggcgggtctgtcttcaagtcaggagtggacgccttgatccacacttgagagaagag
gccagatcagcacccyggctggtagtgatngcagagggcacctgtgcagatctgtgtgc
gcactggaaatctctaggctgaaggcyagagcaaatggcargtgttagtccttgggg
angagagacaganggtgagaaagcaagacagaggtgagagtgacatgtcaagtggta
gattgccttaaaagaaagctaaaaaaagaaagatcggggcgaaacttctttaggggt
aatgctgcagcgtgttaactgactgaccagcgtccatatcttggacccttccccggg
tgaaaaagcccccttcattcctccagcgtcccccaagggtgcttagcaataccgggtgct
tttctgccgcaagtgagttaccaaa

FIG 29A

70/126

...MSAILKVGHHCWLPVTSAVNPQRMLRPPPTAVFNCACCCLWGQMLMNTYRVVQ
LPEEAKGLVPPEILQKYHGFYSSLFALVVRQPRSLQHLCRCALRSHLEGCLPHALPRLP
LPPRMLREFLQLDFFEDLLY*

FIG 29B

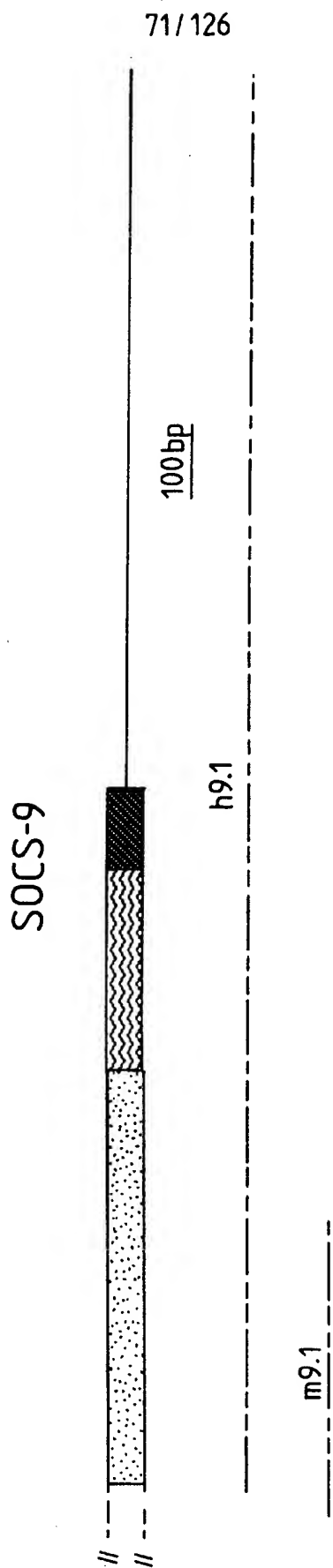


FIG 30

72/126

GTGGGGCGTCATGACCTCCTCTAGGGCTCTGCAACATGACTCCTGTGGTGCAAA
TCAACAAATTGTTCACTGATGAATCCACAAGGATCTCTGGGCTACAACCAGGTCCTG
GTCCACATGACTGTGTCCTTCGGAGAGGACCACTCGCCCCCGGCAGGTACGGCTGA
CACCTCCATGGGAGAAAGACGTATCCAGGCAGCAGCTGCGCGGCCCTTCAAGAGGGCAC
ATCCCGTCATCTAAAGGCACGGTGTAAGGTAGTCTGAGACATGAGTCCGATTA
CTACAGGCACGTGTTCCACAGGTGGAGGCTCAGGTCCCCGGGTGAGCTGGGGCTGCA
GCGGACTCAGGGCGCGCTCTGGCTGCAGGTCTCGCAGCTCCCTGGGCTGTAGCTCC
CGCAGATCCTTGCGCACACCGTTGACTGGT

FIG 31

73/126

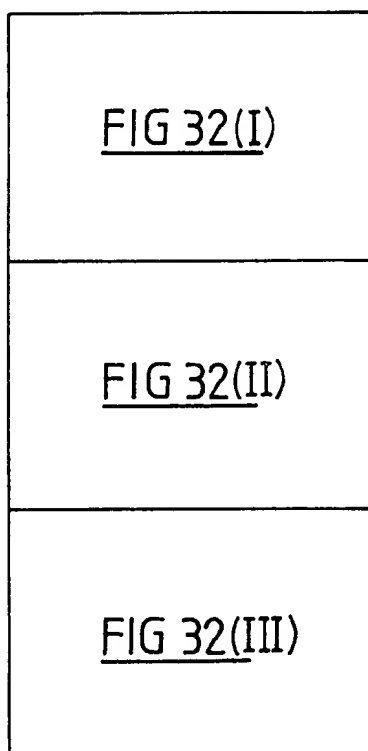


FIG 32

74/126

TTAATAGTACCTACATAGTAGAAAAATTATACTCCACTTTAAAAACAATGTTTCTTCTTC
TATTCAAATCAATTTAAAACTTTTATAAAACATTAATGTTGCAAGAGAAATCCAGTCCA
TTTATGAAAAATTAGTTGACAAATCAAGTTCACCCCAAGAAAAATGTTGACTAAGCTAAAGA
AATCACAGATAAAACATTTTACCAAAAGGATAGGTAACACACAAAAAATGCTATCAC
AGGAAGCTNATGATCATCTAATAATTTCTTTAATAATAATTTCTAGTTCCATAGGTTTTC
ATGTTATGCCAATTTGTACCCGAGTTTAATTACAGAAAAGGCAACAATTTCTAAATTG
GTGGTATACATTTCTTTACAAATTTTAAATGTAAGGCCATTTATTAAAAATAGACAAAC
TAGAAGATGAAAAACGGAAGCAACAGAAAAATTCAACTTTTTCACAAACCAAAAGAAATTAG
CACAACTTAGAAAAATAATTAGAAAAAAGTGTGTTAAAAAGATATGTTGCAGATCTCC
GTTCCATTACCCAAGATATGTCAATTTCACGATTCCTAAATAAATCTTTTAAAGTAAG
AGATTAAAAAATCATCTTCAGTGTATATGTAATAATCCCGTGGTTTATCACAACAGGTAT
GTTTATTCAACACTGKCTTTGGAAANTGGACCATTTAAAAAGGACATGGCAATTTCCAT
TCGTGTTAAGTTTCATTCAACCTTTACTTAGGGGTGRTATTACCACATGAAATGNTGCT
TTTAATGCATAAAAAATCACAGTGGATTAGCCAGCAAAAGGACTGGCGGGGGGGCA
TTGAGGGAGAAATTGATAAATTCACATTTGTGATTATTCTGCACATTGATGAAACATAATT
CACACCTCTAAAAACCTCAAGACTTCCCCTTTTAAAGAACCACAAAATAAACCCAAAGACA
CCTTGCTGACACTTCCCCACCCCTAAACAACTGATGACTCTTTTACACATAAAACTG
AAATAGTTATGGCAGCAAAAGATTTTGATGGCAATGAAAGTTTGTAACACTGTATTTC

FIG 32(I)

75/126

ATCTCTGTCTTATTCCCAAAGTGCAAGATGCAGGGTTCTCAATCTTTCAGTAGTGC
TTCTCCTGTAAATAATCCTTCAATTTTGTGGCAAGGAGTTTCTGAATTAAGTCTA
TTCTGGTATACTGACGTATAACAAAACGACACAGGTACTGCAACGAGCGCACCTSSAT
GAACNCCGRGAACACTGGSTTGGYCAAGTTCTNGACRRGGKAAGKTGCAGATTCCAG
GCAGCYGAGACCTTGAATAACAAAAGCTCCCATTTTCAGAGTCCCTGATTGAATGCT
CCAAATTAGATCAACTATGGACGTATGTCCCTTCCACATCNGGCTGTTCATAAAAAGCTAA
ACCTACCATTTGAGTGCTCAATTCTAGTGTGAAGTGTTTTACCATGGGAGCGAAAGTC
ACAGCTTAAAAGGTAAACGGTCGTCAGAACTGTCCCGAACAAGAAAAGAACCATCTGGC
ACGTTTGCTAGCTTCCCTTCTGCCCTCCCAACGTGTGATTGGTCCCCAGTACCATCCTT
GCTTTGCAAGTTTTTTCAGCTCCTCTGTAAAGCTTGTCAACAACCATGGGACCACACT
TTGCCACTGAGTCATAAACTCTTGCAACCCAGGAGCAGAGTTCGGATCAAAAATTCAAA
TGACAGCGCATAAAGTTTNCAGCCACGTGGGGCTTCTGTSCCAGTGAGTCCACTGAAA
GTTCCCTTTGGGATTTGGATTATTCTCTGCATTGGAGNTAACCAATGGTGAAGATTGG
AGGGACATCCATCGTGAACCCGCTCTCCGGGGTTCTGCAACATGACTCCCCTGGTGCC
AATCAACAAGCCATTACCCGGACTGATCCACGAAGATCTCTGGGGCGACAACCTAGGTC
CTGGTCTACCTGACTCTCATCTCGGGGAAAGCGGCCCTCCCACTTGAGGAGGAACC
GCAGAGACTTCCATGGGAGAAGAGCTGTCCAGACAATAAGTCCGTGATCCTTCCAAAG
GATACATCCCCCTCATCTAAAGGCACAGTATACTGAATGTAGTCTCTGAGGCATAAGTCC

FIG 32(II)

76/126

AAATAACGACAGGCACATGTTTCATCCAGGTGAAGATGCAGGTCTCCATTATGAGAAAGCC
GAGCTCTTCAGTGAATTGGCTTGCTCCTGGCACGTTGGTCTCAGACTGGAGGTCGT

FIG 32(III)

77/126

SOCS-10

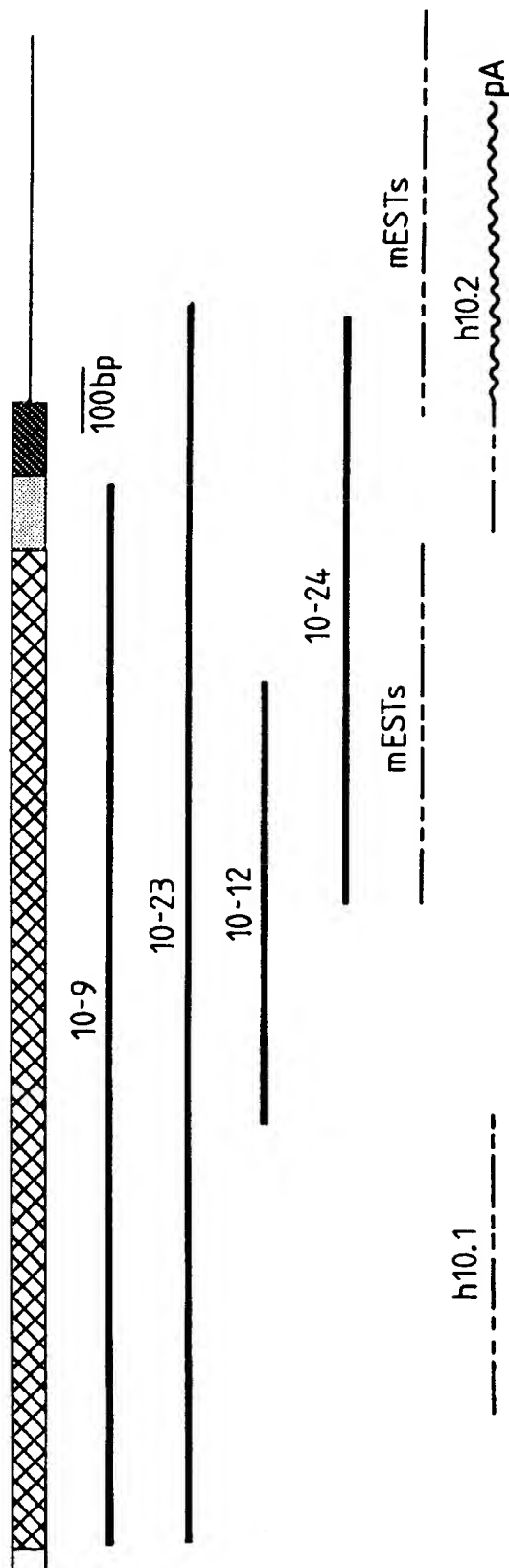


FIG 33

78/126

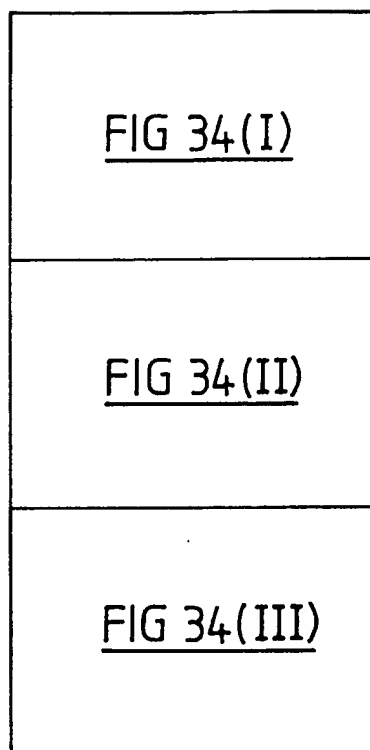


FIG 34

79/126

GGCACGAGGCTGTGTCCAGCACACAGAGAGGGCCCGGCCATCTGCTTTGGTTCAGAGC
CCTGTGTCTGTCTACTTAGACTCTTCCTCCCGGCTCGAGCTCACCCCTCCATCCT
CCTTACTGGCTCCAGCATGACTCGCTTCTCTTATGCAGAGTACTTTGCTCTGTTTCAC
TCTGGCTCTGCACCTTCCAGGTCCCCCTTCGTCTCCCGAGAACCCACCGGCCCGGCAC
CCCTGGGTCTGTTCCAGGGTCAATGCAGAAGTATAGCAGCAACCTGTTCAAGACCTC
CCAGATGGCGGCTATGGACCCCGTGTGAAGGCCATCAAGGAAGGGATGAAGAGGCC
TTGAAGATCATGATCCAGGATGGGAAGAACTTGCAGAGCCCCAACAGGAGGGCTGGC
TGCCGCTCCACGAGGCTGCCCTACTATGGCCAGCTGGGCTGCCCTGAAAGTCTCTGCAGCA
AGCCTACCCAGGACCATTGACCAACGCACACTGCAGGAAGAGACAGCATATATACCTG
GCCACATGCAGAGAACACCTGGATTGCCCTCCTGTCTGCTCCAGCGGGGCGCAGAGC
CTGACATCTTAACAAATCCAGGGAGACTCCACTTTACAAAGCCTGTGAGCGCAAGAA
CGCGGAGGCGGTGAGGATATTGGTGCGATACAACGCAGACGCCAACCCCGCTGTAAC
AGGGGCTGGACCGCACTGCACGAGTCTGTCTCCCGCAATGACCTGGAGGTCATGGAGA
TCCTAGTGAGTGGCGGGCCAAAGTGGAGGCCAAGAATGTCTACAGCATCACCCCTTT
GTTTGTGGCTGCCAGAGTGGCAGCTGGAGGCCCTGAGGTTCCCTGGCCAAGCATGGT
GCAGACATCAACACGCAGGCCAGTGACAGTGCAATCAGCCCTCTACGAGGCCAGCAAGA
ATGAGCATGAAGACGTGGTAGAGTTTCTCTCTCAGGGCGCCGATGCTAACAAAGC
CAACAAGGACGGCTGCTCCCCCTGCATGTTGCCCTCCAAGAAGGGCAACTATAGAATA

FIG 34(I)

80 / 126

GTGCAGATGCTGCTGCCGTGACCAGCCGCACGGCGTGGCCGTAGCGGCATCAGCC
CGCTGCACCTAGCGGCCGAGCGAACACGACGCGGTGCTGGAGGCGCTGCTGGCCGC
GGCTTCGACGTGAACGCACCTCTGGCTCCCGAGCGCGCCCGCTCTACGAGGACCGC
CGCAGTTCTGCGCTCTACTTCGCTGTGGTCAACAACAATGTGTACGCCACCGAGCTGT
TGCTGCTGGCGGCGGACCCCAACCGCGATGTATCAGCCCTCTGCTCGTGGCCAT
CCGCCACGGCTGCCCTGCGCACCATGCACTGCTGTGGACCATGGCGCCAACATCGAC
GCCTACATCGCCACTCACCCCAACCGCCCTTCCAGCCACCATCATGTTTGCCCATGAAGT
GCCTGTCGTTACTCAAGTTCTTATGGACCTCGGCTGGGATGGCGAGCCCTGCTTCTC
CTGCCGTGACGGCAACGGGCCGCACCAACCGCCCGCGACCTGGCCGCTTCCACGACG
CACCCGTGGACGACAAGGCACCTAGCGTGGTGCACTTCTGTGAGTTCCCTGTGCGCC
GGAAGTGAGCCGCTGGGCGGACCCATCATCGATGTCTCTGACTATGTGGGCAAC
GTGCAGCTGTGCTCCCGCTGAAGGAGCACATCGACAGCTTTGAGGACTGGGCTGTCA
TCAAGGAGAAGCAGAACCTCCGAGACCTCTGGCTCACCTCTGCCGGCTGCGGGTTCCG
GAAGGCCATAGGAAAAATACCGGATAAACTCCTGGACACACTGCCGCTTCCCCGGCAGG
CTAATCAGATACTTGAAATATGAGAAATACACAGTAaccagcctggagaggagatgtg9
ccttcagactgtttccg9gagccccccaggtggcctgcatccaggacccccctgg9gtca
gaacaggtgtgaccttgctgggttcttctgtggagcttcaacccaaagtgaagaacctgat
gtgggagtg9gacgtggaacccctctgctttcacactgtcagcggatcgcagaccccgctc

FIG 34(II)

81/126

tgcttctggccatagccagagaccttcaacctggggccaggaggagagctggtcttgggc
aaggtagccaggcaggaaatcctggccttaagctggagaactttaggaatccctcac
tggacctcagcttctcaggctgcgaggagacgcccagcccagctattttatttcwgcg
tgacacaaacgttgtatcagaaaaaaacacatgggcccagcttattccttag
tagggtaattacttgcatgcngcgcttaagcctactggaaacatgcttccnactat
gcttgagaatcccccttgccactggtaaacgagagccgacgtgcttcaagggttgatattt
tggnctgcccccttggcgttccgcgggttgntccgacngtaattgaccccgtgtttt
gtcacttgcgagtggtccgactattggggggttctgggtgtccccaaaattgtgggt
ggtgtgcggacgccacgagaaagtgttcatgggcgataatcattactgngagaaatgta
gagcggcgggtttacgaataaataatttttaagccgccttccccaaa

FIG 34(III)

82/126

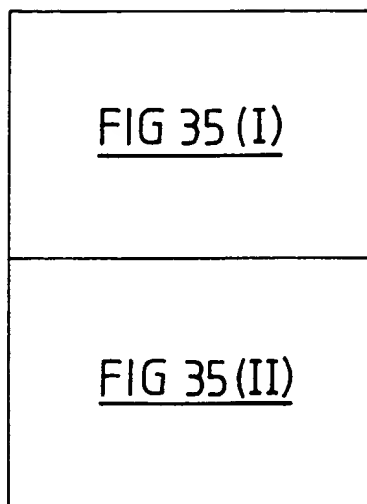


FIG 35

83/126

h10.1

CCTCCTGAGAGTTCGCCGGCCCGGCCCAATGGGnTTGTTCCAAGGGGTCAATGCAGAA
ATACAGCAGAGCTTGTTCAGACCTCCAGCTGGCGCCTGCGGACCCCTTGATAAAG
GCCATCAAGGATGnCGATGAAGAGGCCCTTGAAGACCATGATCAAGGAAGGAAGAATC
TCGCAGAGCCCAACAAGAGGGCTGGCTGCCGTGCACGAGGCCGCATATAATGGCCA
GGTGGGCTGCCCTGAAAGTCTGCAGCGAGCGTACCCAGGGACCATCGACCCAGCGCACC
CTGCAGGAGGAACAGCCGTTTACTTGGCAACGTGCAGGGGCCACCTGGACTGTCTCC
TGTCACCTGCTCCAAGCAGGGGCAGAGCGGGACATCTCCAACAATCCCCGAGAGAnACC
GCTCTACAAGCCCTGTGAGCGCAAGAACGCGGAAGCCGTGAAGATTCTTGGTGCAGCA
CAACGCAGACACCAACAACGCTGCAACCCGGGCTG

FIG 35 (I)

84/126

h10.2

GTGCAGCTCTGCTCGCGGCTGAAGGAACACATCGACAGCTTTGAGGACTGGGCCCGTCA
CAAGGAGAAGCAGAACCTCCAAGACCCTCTGGCTCACCTTTGCCGACTGCGGGTTCGAA
AGGCCATTGGGAAATACCGTATATAAACTCCTAGACACCTTGCCCGCTCCAGGCAGGCTG
ATTAGATACCTGAAATACGAGAACACCCAGTAACCTGGGGCCACGGGGAGAGAGAGTAG
CCCCTCAGACTCTTCTTACTAAGTCTCAGGACGTCGGGTCTCCCAACTCCAAGGGGACC
TGGTGACAGACGAGGCTGCAGGCTGCCCTCCCTCTCAGCCTGGACAGCTACCAGGATCTC
ACTGGGCTCTCAGGGCCAGAGCTTTGGCCAGAGCAGAGAACAGAAATGTGTCAAGGAGAA
GAATCATTTGTTTACAAACTGATGAGCAGATCCAGACCCTTCTCTACCTTCAGGAATGG
CAGAAACCTCTATTCTGGGGCCAGGCAGAGCTTGAGGTGTTCTGGGGAAGTGGTGC
TCAGAGCCCTTCCCTGTGCCCCCTCCACTTGTCTTGGAATACTACCACTTGACTTCAGAG
CTTCTCTCCAAGACTAAGATGAAGACGTGGCCCCAAGGTAGGGGGTAGGGGGAGCCCTG
GGTCTTGAGGGCCTTTGTTAAGTATATAATAATAAATGTTACACATGTGAAAAA

FIG 35 (II)

85/126

TTGGAGAAAGTGTGGTTGGTATTGGGGGGCCAAATGAATTGGGAAGATGCAGAGATGAAGC
TGAAAGGGAAACCAGATGGTTCTTCCCTGGTACGAGACAGTTCTGATCCTCGTTACAT
CCTGAGCCTCAGTTTCCGATCACAGGGTATCACCCACACACTAGAAATGGAGCACTAC
AGAGGAACCTTCAGCCTGTGGTGTCATCCCAAGTTTGAGGACCGCTGTCAATCTGTTG
TAGAGTTTATTAAAGAGAGCCATTATGCACTCCAAGAAATGGAAAGTTTCTCTATTCTTT
AAGATCCAGGGTTCAGGACTGCCACCAACTCCTGTCCAGCTGCTCTATCCAGTGTC
CGATTTCAGCAATGTCAAATCCCTCCAGCACCTTTTGCAGATTCCGGATACGACAGCTCG
TCAGGATAGATCACATCCCAGATCTCCCACCTGCCCTAAACCTCTGATCTCTTATATCCG
AAAGTTCTACTACTATGATCCTCAGGAAGAGGTATACCTGTCTCTAAAGGAGCGCAGCGT
CAGTTTCCAAACAGAAAGCAAGAGGTGGAACCTCCACGTAGCGAGGGCTCCCTGCTG
GTCACCAACCAAGGCAATTGGTTGCCAAGCTCCAGCTTTGAagaaccaaataagcta
ccatgaaaagaaggaagaaagtgaagggaacaggaagggtgggattctctgtgcagaga
cttttggttccccacgcaagccctggggcttggaagaagcacatgaccgtactctgcgt
ggggctccacctcacaccccccctgggcatcttaggactggaggggctccttgga
actggaagaagtctcaactgtttcttttcca

FIG 36A

86/126

...LEKCGWYWGPMNWEDAEMKLKGKPDGSLVRDSSDPRIYILSLSFRSQGI THHTR
MEHYRGTFSLWCHPKFEDRCQSVVEFIKRAIMHSGKNGKFLYFLRSRVPGLPPTPVQLL
YPVSRFSNVKSLQHLCRFRIRQLVRIIDHI~~PDLP~~PKPLISYIRKFYYYDPQEEVYLSL
KEAQRQFPNRSKRWNPPRSEGLPAGHHQGHVLVAKLQL*

FIG 36B

87/126

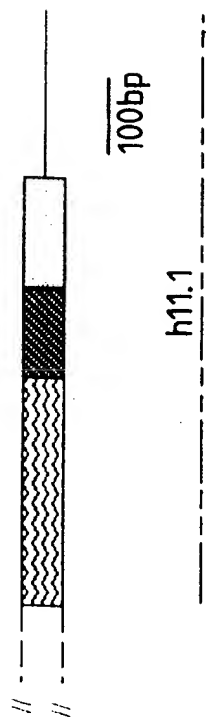
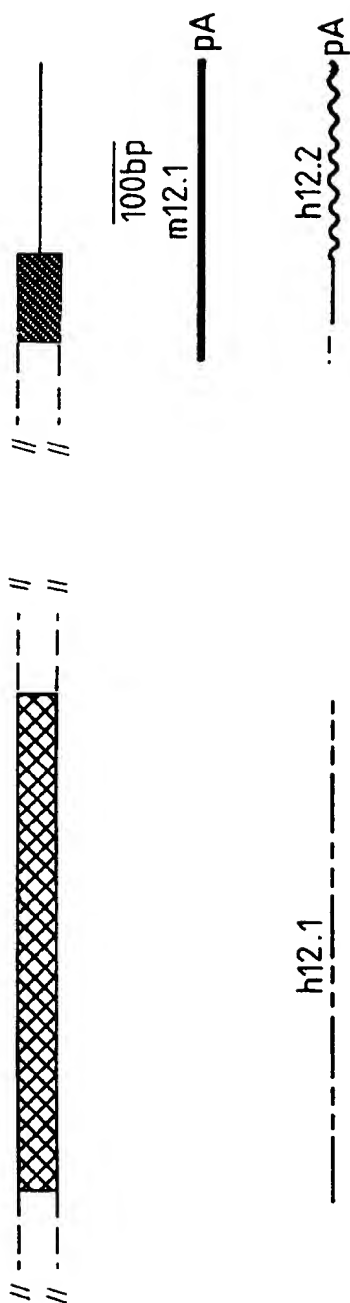


FIG 37

88/126

FIG 38



89/126

GTTCGAAGCCTAACCCATCTTTGTCTGTTTGGAATTCGGGCCAGTCTAAAAGCAGAGC
ACCTTCACTCTGACATTTTCATCCATCAGTTGCCACTTCCAGAAAGTCTGCAGAACTA
TTTGCTCTATGAAGAGAGGTTTAAAGAAATGAATGAGATTCTAGAACCCAGCAGCTAATCAG
GATGGAGAAACCAAGCAAGGCCACCTGACacaggtcctttaattctgttttagtcacaaa
agacggcttggtgactgttttggtatcgaatgtccatgttttacagtttgctt
ttcccagtttggtgtctttcccgaatattgtgaaccttatccatcttgccttactcagtt
ttatttctagtgactttgtgtgtattattgttttacctgaccatttttctactttat
tctgctaataaactgtaattctgaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

FIG 39

90/126

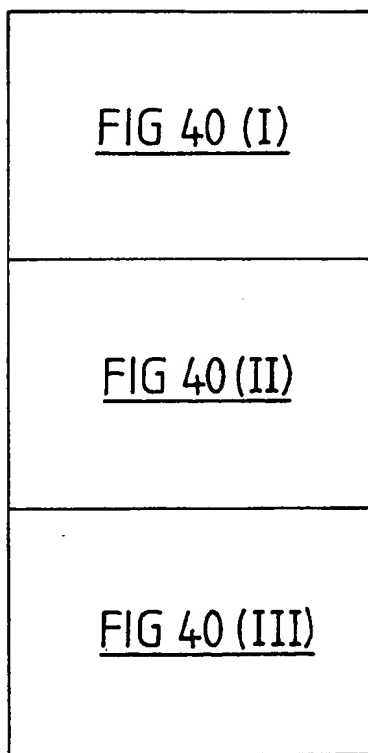


FIG 40

91/126

h12.1

GGGATCGAAAGCGGGGCTTCTGGACGCAGCTCTGGAGACGCGGCTCGGACCAGC
CATTCGGGTAGAAAGTGGCAGCACGGCAGACTGGTCAAAACAATGGATTTACAGAG
GCTTACGGGACACGTGCTCTACAGTTGGACTTGCTGCCAGGGAAGCAATGTTAAAG
TCTTAAGGAAACTGCTCAAAAAGGGCCGAAGTGTCTGATGTTGCTGATAACAGGGGATG
GATGCCAATTCATGAAGCAGCTTATCACAACTCTGTAGAATGTTGCAAAATGTTAATT
AATGCAGATTCACTGAAAACACATTAAGATGAAGACCTTTGAAGGTTTCTGTGCTT
TGCACTCTCGCTGCAAGTCAAGGACATTGGAAAAATCGTACAGATTCTTTTAGAAGCTGG
GGCAGATCCTAATGCAACTACTTTAGAAGAAACGACACCATTTGTTTGTAGCTGTTGAA
AATGGACAGATAGATGTGTTAAGGCTGTGCTTCAACACGGAGCAAAATGTTAATGGAT
CCCATTCTATGTGTGGATGGAACCTCCTTGCAACCAGGCTTCTTTTCAGGAAAAATGCTGA
GATCATAAAAATTGCTTCTTAGAAAAAGGAGCAAAACAAGGAATGCCAGGATGACTTTTGGGA
ATCACACCTTTATTGTGGCTGCTCAGTATGGCCAAGCTAGAAAAGCTTTGAAGCATAC
TTATTTTCATCCGGGTGCAAAATGTCAATTGTCAAGCCTTGGACAAAGCTACC

FIG 40 (I)

92/126

h12.2

CACAAATGGGACCATACAAAAATCTTGGNACTTGTTAATAACCACCTTNACTAACCGGG
ACCTGTGACACTGGGNCTAAACAAAGTAAGTCCCTGTCTTACTCAGNCAGTGTTTGGGG
GACATGAAGGATTGCCCTAGNAAATATTACTCCGGAAATGGTCTACAGCCCAGNACGCCCC
AGGCGTGCCCTTGTTTTGGGATTCAGTTCTCCTGTGTGCATGGCTTTCCAAAAGGAGGT
GGAGCTGTRAGTTCTTTGGAAATTGTGAACATTCTTTTGAAATATGGAGCCCAGATAAA
TGAACCTTCATTTGGCATACTGCCCTGAAGTACGAGAAGTTTTCGATATTTCGCTACTTT
TTGAGGAAAGGTTGCTCATTTGGGACCATTGGAAACCATAATATGAATTTGTAAATCATG
CAATTAAAGCACAAAGCAAAATATAAGGAGTGGTTGCCACATCTTCTGGTTGCTGGATT
TGACCCACTGATTTCTACTGTGCAATTCTTTGGATTGACTCAGTCAGCATTGACACCCCTT
ATCTTCACCTTTGGAGTTTACTAAATTGGAAGACACACTTGCCACCAGCTGTTGAAAGGATGC
TCTCTGCTCGTGCCCTCAAACGCTTGGATTCTACAGCAACATAATTGCCCACTGTTCCCAT
CCCTGACCCCATCTTTGTCTGTTTGGAAATTTCGGTCCAGTCTAAAAATCAGAACGTCCTACG
GTCTGACAGTTATATTAGTCAGCTGCCACTTCCCAGAAGCCTACATAATTATTGCTC
TATGAAGACGTTCTGAGGATGTATGAAGTTCCAGAACTGGCAGCTATTCAAGATGGAT
AAATCAGTGAAACTACTTAACACAGCTAATTTTTTCTCTGTGAAAAATCATCGAGACAA

FIG 40 (II)

93/126

AAGAGCCACAGAGTACAAGTTTTTATGATTTTATAGTCAAAAGATGATTATTGATTGT
CAGATAGGTTAGGTTTGGGGGGCCAGTAGTTCAGTGAGAAATGTTTATGTTTACAACT
AGCCTTCCCAGTAAAAAATAAAAAAATAAAAAA

FIG 40 (III)

94/126

S0CS-13

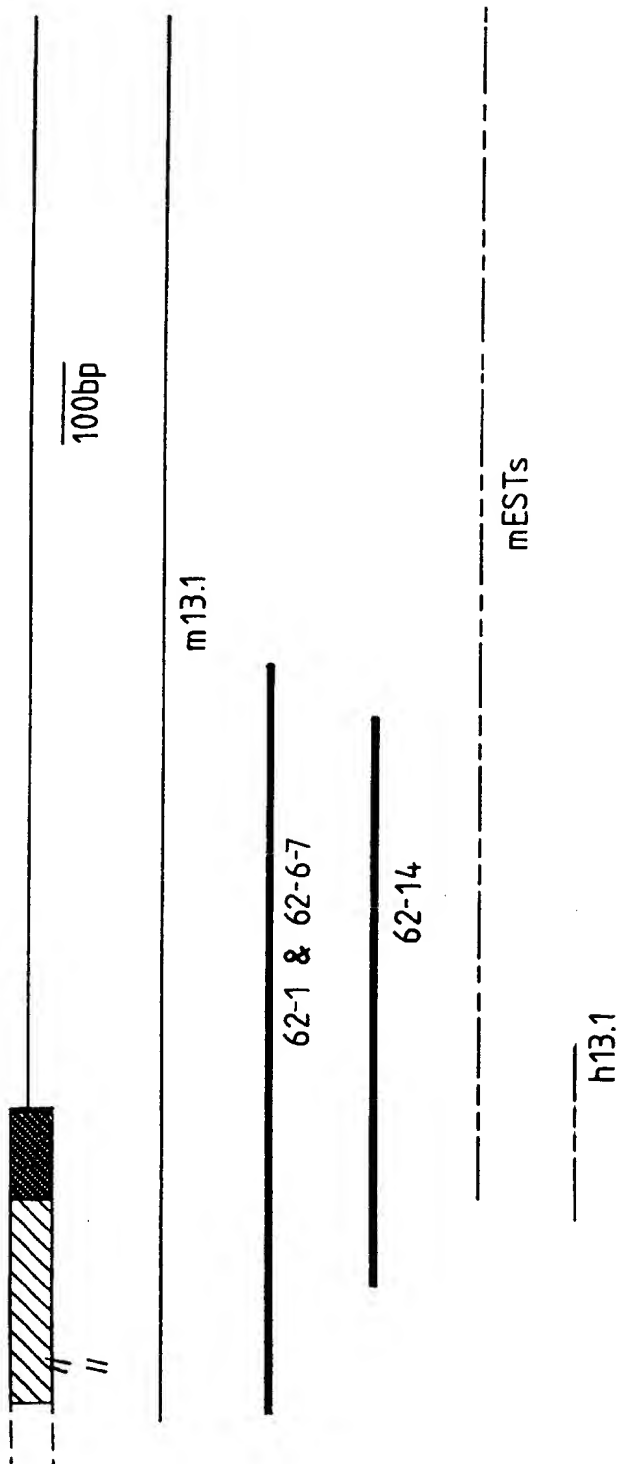


FIG 41

95/126

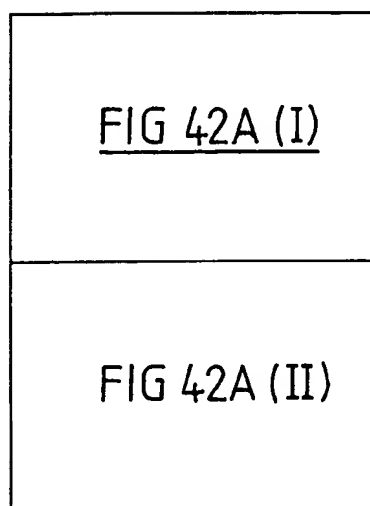


FIG 42A

96/126

CGGGGGCTGGGACCTGGGGCGTAACCGTCTCTACCACGCGCAAGAACAGCCAAG
TAAACATACCCAGCCTTTCTGGAGCCGGACGAGACATTGATGTCCCTGACTCCTTT
TTCCGTGGCCCTGGACATGRATGATGGACCTTAAGTTTCATCGTGGATGGACAGTACA
TGGGAGTGGCTTTCGGGACTCAAGGGTAAAGCTGTATCCTGTAGTAGTGCCCGT
CTGGGGCCACTGTGAGATCCGCATGCGCTACTTGAACGGACTTGATCCTGAGCCCCCTG
CCACTCATGGACCTGTGCCGGCGTTTCGGTGCGCCCTAGCGCTGGGAAAAGAGCGCCTGG
GTGCCATCCCCGCTCTGCCGCTACCTGCCCTCCCTCAAAGCCTACCTCCTTACCAGTG
Atccacatcccaggaccgccatacagacagccatctggtgccaaartcactgagccccgtt
ggggtccgcgacccctgcgcctgggagtggaygccacccctcagccatgggcagacgtg
ccccctcatcctaccggctgcctctgctggggaaacctatgccaaaggacttctccct
tcccaaacctggctgaagcagcagcaccagggcccttccctgaaccagatgcagagaa
taaaactatgaaaaacctctctcagggccttctgctctcagggtggagtgggctgcccc
cactctctgcagagagaggctacaccacctggggggtcctgggaggtaagactagta
ggagggtccagggtgartccaaaagcagggaatggccaggamcaggccatacagatga
agctcaggatgtcacataccatggacamtgagacagaaacccagggttggamttccctt
gggccaacgagtgccagctttaatgtcagctgcmgggtcctgtggcctgtatttatt
ctttaaacagtagcaaaaggccatttatttccacttagaaaggaaaccttggtggg
tggyttccctcgatgtgcttccccccacctccctggaatgtgtgtgccacacctgtcc

FIG 42A(I)

97/126

ttgtcccaggcaggactgtggcacatgagctggtgtgcacagatacacgtatgtcgt
cgtgcatgacccctgactagtccctaagtagccctgcaccaagcaccagagcagacccc
caagagagggcccggtgcaagtccccatgtcccagggtccctgcttctgttgccttggga
ctcatacaccggcacacgtgtttcagccctcttgacttccatgagcttcgaaatttggcc
ccgattcttctgatatctcccatggcatcctccaaagctctgggcctggaggggcat
taggacacatgggaatgagtgggtctccagccctgggaaaagccactggcaaggcagg
attagaaagaccaagagcaggggtggggcgccatgaagcctgtatgcctcttcaggctca
agaccccgccacacacccactcaagcctcagaagtgggtgtgtagggcagccccaggag
aggaaatgcctgtcctagcagcacgtacatggagcaccccccacatgtgtccagccctct
ggctgtttctcttgctctagaatacaactccctacatctgggaatgtagccatttggtag
aggacttgcctagcctgcagggaagctcacgttccatccccctgcaccaaggagaatcaa
agctcaggagggtgaggcaggaggattgtctgtcagtgggtgtacagaggtcatggccat
cctgggctatatataaaccttgtcctttaagaaaaaagaaaaatcaacttccattga
atctgagttctgctcatcttctgcacagggtacaaatagatgacttkatttgttgaaaaat
gkttaatatatttacmtatatatatatttgtgaagaagcatt

FIG 42A(II)

98/126

...GGWDLGRNRLYHDKGNQPSKTYPAFLPEDETFFIVPDSFFVALDMXDGTLSEFIVD
GQYMGVAFRGLKGKKLYPVVSAVWGHCEIRMRYLNGLDPEPLPLMDLCRRSVRLALGK
ERLGAIPALPLPASLKAYLLYQ*

FIG 42B

99/126

AAGGGTAAAAAACTGTATCCTGTAGTGAGTGCCGTCCTGGGGCCACTGTNAGATCCGAA
TGCGCTACTTGAACGGACTCGATCCCGAGACNTGCCGCTCATGGATTGTGCGCTCGC
TCGGTGCGCCCTGGCCCTGGGGAGGGAGCCCTGGGGGAGAACCAACNACCTGCCCGCTG
CCGGCTTCCCCTCAAGGCCCTACCTCCTACCAAGTGACGTTCCGCCATCATACCGCCAGC
GCGACAGCCACCTGGTGCCAACTCACTGAGCCGCGCTG

FIG 43

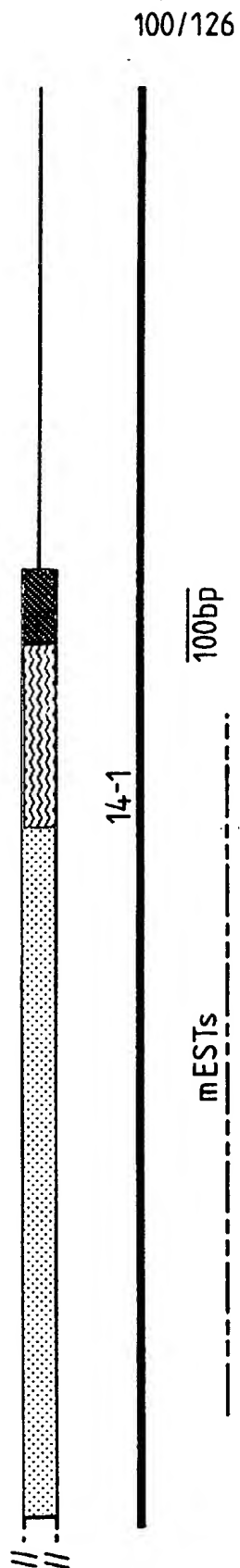


FIG 44

101/126

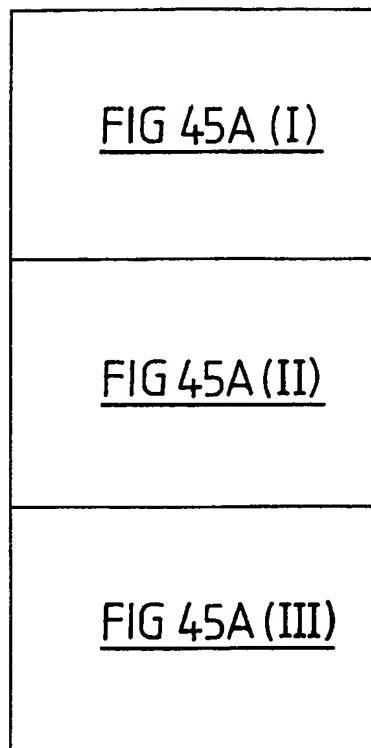


FIG 45A

102/126

...AAGTGGCGGGTCCCTGGAGAGCAGCGGAGGCAGCGCAAGTCTGACTCTGG
GCTGACCGTGGAGCCGGGGGGGGCTGACAGCCAGGCCCTCCGCCCTGGCGGAGCCGC
ACGAGGAGCGGAGTGGCCGGCCCTCTCTTCCGCGCTTGAGCGAGCGCCGGTGATGG
CGGTGGTGATGGCGGAGCGCTCGGACAGCTCCGCTTGAGCTGAGCTCGGAGAGATC
CGTCCAGAAAGTGCCCCAGAAAGAACTTCCTCTTAGAAAGCTGAAAAACACARTATTT
ATAACACTGGAAATTGTAAAGAAATTGTTTAAAAATGGCTGAAAAACAATAGTAAAAATG
TAGATGTACGGCCTAAACAAGTCGGAGTCGAAGTGCTGACAGGAAGGATGGTTATGT
GTGGAGTGGAAGAGTTGTCTTGGTCCAAAAAGAGTGAGAGTTGTTCTGAACTCTGAA
GCCATAGGTA CTGTTGAGAAATGTTGAAATTCCCTCTAAGAAAGCCAAAGCAGCCTTA
GCTGTTCCGTCATTGAGTTGGACTTAGATCATTCCTGTGGGCATAGATTTT TAGGCCG
ATCCCTTAAACAGAAACTGCAAGATGCGGTGGGCAGTGTTTCCCAATAAAGAAATTGT
AGTGGCCGACACTCTCCAGGGCTTCCATCTAAAGAAAGATTCTATATCAGTGAACCTCA
TGTTAGATAAGTGCCCTTTCCCACTCGCTCAGATTTAGCCCTTTAGGTGGCATTTTAT
TAAACGACACACTGTTCCTATGAGTCCCAACTCAGATGAATGGGTGAGTGCAGACCTG
TCTGAGAGGAAACTGAGAGATGCTCAGCTGAAACGAAAGAAACACAGAAGATGACATAC
CCTGTTTCTCACATACCAATGGCCAGCCCTTGTGTCAATACTGCCAACAGTGCTTCGTG
TACAGGTGTCACATAA ACTGGTTCTATGATGAAC TTGGTCACAAACACAGCATAGAA
GACAGTGACATGGATTTCAGAGGATGAAATTATAACGCTGTGCACAAGCTCCAGAAAAA

FIG 45A(I)

103/126

GGAA TAAGCC CAGTGGGAAATGGAAGAGGAGATCCTGCAGTTGGAGGCACCTCCTAA
GTTCCACACCCAGATCGACTACGTCCACTGCCCTTGTTCCAGACCTCCTTCAGATCAGT
AACAAATCCGTGCTACTGGGGTGTCATGGACAAATATGCAGCCGAAGCTCTGCTGGAAG
GAAAGCCAGAGGGCACCTTTTACTTCGAGATTCAGCGCAGGAAGATTATTATTCTC
TGTTAGTTTAGACGCTACAGTCGTTCTCTCATGCTAGAAATTGAGCAGTGGAATCAT
AACTTTAGCTTTGATGCCCATGATCCTTGCTGCTTCCATTCTCCTGATATTACTGGGC
TCCTGGAAACACTATAAGGACCCAGTGCCCTGTATGTTCTTTGAGCCGCTCTTGTCAC
TCCCTTAATCCGGACGTTCCCCCTTTTCCCTTGACAGCATATTGACAGAACGGTTATTGT
AATTGTACGACTTACGATGGCATCGATGCCCTTCCCATTCCTTCGCCCTATGAAATTGT
ATCTGAAGGAATACCATATAAATCAAAGTTAGTTACTCAGGATTGATGTGCCAGA
GCAGCAGTGATgcgagaggtagaatgtcgacctgcatacatatttcatataat
tttatttttcttatgcctctttgaatttttgtacaaaggcagttgaatcaataaaac
tgtgcccctaagttttaattccagatcaatttattttttatgatacacttgtttatat
attttaagcagggtttgggtttgtttttaccataataaattacatatgtggtccaggc
atatttacaatttcaaggcattgcatatacatatttgaatatctgtattttttaataa
tcttttgttcttcctatgtgtgaaaatttttgctaattctatgctatcagtatcttg
tatgaccgaatagttacctattctcttttcatcttgaagattttcagtaagagtggt
gtaatcaatccattataatgtaattgacttttgttaatttgccaataggagtgtaa

FIG 45A(II)

104/126

aacaaaatgatttaaaatgaaacttaatgtattttcattttaaatataactaaacca
agttgtttgttagttattctagccaataagaaaaagagaatgtagcatcctagaggtg
tatttgttctgcagtttggcaggaccgtcagtttagtccaaataaacatccccctcagcg
tggaggcggaatggaaacctgtgctcctttcttacgggaagcttgcaaaagcaaaatagc
agggttacaagcttgagttgtaaggcaactagagttttctctatttaattatagac
tgttgttgcaacctacttagctcttttttggggaactctagttcccagggggaaataacct
cgtgcc

FIG 45A(III)

105/126

....SGGGPWAGGGSGKSDSGLTVEPGRGLTARPPPGGSRTRSGSGRASLPRLSERR
VMAVMAAGARTAPLELSSERSVQVPRRNFLLEKLKNTXFITLEIVKNLFKMAENNS
KNVDVRPKTSRSDRKGYVWSGKKLSWSKKSESCSESEAI GTVENVEIPLRSQER
QLSCSSIELDHSCGHRFLGRSLKQKLQDAVGQCFPIKNCSGRHSPLPSKRKIHIS
ELMLDKCPFPPRSDLAFRWHFIKRHTVPMSPNSDEWVSADLSERKL RDAQLKRRNTED
DIPCFSHNGQPCVITANSASCTGGHITGSMNMLVTNNSIEDSDMDSEDEIITLCTSS
RRNKPRWEMEEEEILQLEAPPKFHTQIDYVHCLVPDLLQISNNPCYWGVMDKYAAEAL
LEGKPEGTFLLRDSAQEDYLFVSFRRYSRSLHARIEQWNHNFSFDAHDPCVFHSPDI
TGLEHYKDPSCMFFEPLLS TPLIRTFPFSLQHI CRTVICNCTTYDGDALPIPSPM
KLYLKEYHYKSKVRLLRIDVPEQQ*

FIG 45B

106/126

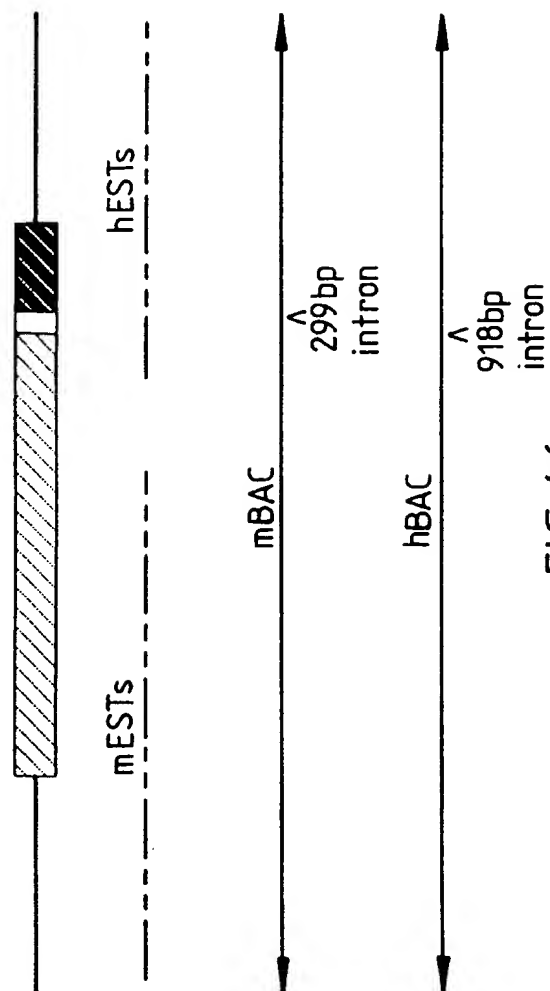


FIG 46

107/126

<u>FIG 47A(I)</u>
<u>FIG 47A(II)</u>
<u>FIG 47A(III)</u>
<u>FIG 47A(IV)</u>
<u>FIG 47A(V)</u>

FIG 47A

FIG 47A(I)

[illegible]

109/126

gactaggctagccttgaactcagagatccgcctgcctctgcctccaagtgtgggat
tataagtggtgcaccaccactgcccagccactttgggatttttgaactgttatcaaga
ggctttcaggagggtcaaaacttcaacaggaacctctccatgataatgtagctaatgatac
aaacgacactcaaaacttaacccttaaaagcacacatccaccagacagcgtgccactc
gtagttccattactcaggaggctgaagcaggaggatgaaggactaaggcttcagcaac
ctaggagccgcaggggacagtagtctcaatccctacattctcctgaacacaggagca
ggagttcaggaaagggtgtcaaggccgcttactgatctttagggcctcagggaatgactag
ctcaggcagagagcaaaagggtctccagtgaggaaagtctacacacacacacacaca
cacacacacacacacacacagaatccaaggcgatgacgtcatcaaaagggttaattc
tagtctgggattggggggagggtggggcacgcagctgtcaggtggctttggaaaaata
aactgctgaagagtctgacgccaggaggtcctgggagggacaaagaggttaccactca
aagagtgtgtccacaaagcatgcgcgcttgtccacgtctggagtcgtcacttattt
ttgcctggattctttgtagccggtgggttctcaaggcggtaagtgggtggccgcgt
ggtctgggaggtgacgatatagggttaatcgtccacagagcccaggggcgagcggggc
gggcgtccgcagccccgcgtggagccggaagcagtggtcgtcagggcgcttctagcc
ttccctatctgtacttccacagaggtctctgcagctagggggacagtgagggtcgggg
gtagggggcccggttagagccagcaaggggacggttcacggtaagggtctgagggaga
gagagctcctgagaaacttggggggcgacacagatatagggtgaaagcagagtgatag

FIG 47A(II)

110 / 126

acctgggatggttagggaccaaggaagaccaggctgggtggcatacaccggtgaac
ggatgggagtcctagggaaagatgatgcgcctaacagtccttctgtctccacaccac
tccagggacgatccggagctcaactttcaaaagcgagacgccccagcaagcctgttt
tgagaaagtcttcagcggtctctctcATGGGCCAGACGGCCCTGGCAAGGGCAGCAG
CAGCACCCCTACCTCGCAGGCTCTGTACTCGGACTTCTCTCCTCCGAGGGCTTGGAG
GAGCTCCTGTCTCCCCCTCCTGACCTGGTTGCCCAACGGCACCCACGGCTGGAACC
CCAAGGATTGCTCCGAGAACATCGATGTCAAGGAAGGGGTCTGTGCTTTGAGCGGCG
CCCTGTGGCCAGAGCACTGATGGAGTCCGGGGAACGGGGCTATTCGAGAGGTCTG
CAGGCTGGGAGATCAGCTGGCCCCCTGGAGCAAAGGGGCACACGCCGTGGTGGGCG
TGGCCACCGCCCTCGCCCCGCTGCAGGCTGACCACTATGCGGCGCTTTTGGGCAGCAA
CAGCGAGTCCTGGGCTGGGATATTGGGCGGGGAAATTGTAATCATCAGAGTAAGGGC
CTCGAGGCCCCAGTATCCAGCTGGACCTCAGGGTGAGCAGCTAGTGGTGCCAGAGA
GACTGCTGGTGTCTGGACATGGAGGAGGGGACTCTTGGCTACTCTATTGGGGGCAC
GTACCTGGGACCAGCCTTCCGTGGACTGAAGGGGAGGACCCCTCTATCCCTCTGTAAGT
GCTGTTTGGGGCCAGTGCCAGGTCCGCATCCGCTACATGGGCGAAAGAGAGgtgaga
tacggactagggtgtgggagatcactactcttggcaatggtttgggctggaaactcat
ggttggagcacaggaagtaggcttcttgtcactttggcctgtcacttagatggccttg
gatctagcttcactcccaatccctatttgatgtgatgcacaaatcagagcccttggg

FIG 47A(III)

111/126

tctccctcagctgaggtggcggtggaatggaggaaggaagggtgcctgagcagg
atctcaagttcaaggatgcctggagttgcttacttaccttgtcttcttctctccg
cagTGGAGGAACCAATCCCTTCTGCACCTGAGCCGCTGTGTGTGGCCATGCTCT
GGGGACACCCGGTGGTCAAAATATCCACTCTGCTTGTCCCCCTGCCATGAAGCGC
TATCTGCTCTACAAATGAcccagtagtacaggggtgctggcacccctaccgtggggac
aggtggagaggcaccgcctggcctagacaaactttaaaaagctggtgaagctggggggg
ggggctggacccttcacctccccctctcacaggagcaagacatatagaaatgatat
taaacaccatggcagcctgggacaaaagggttttgaagtaaaaaatgagatgtatg
tcacaaacctgtttcatattgtttttgttttttacactccccaccagggtta
gagccccatcactgtcttaaggaaattatgacaaacccacaagctcagggccagggtgtt
tattcccttacatgtaggatggttcacaaacacaatacacaggggctttggcacctgg
ggaggggactatcccaggcctcttagggtctcatgtataccggaattcagacccgaaa
gctctgaatttctgcatcagacatccagtagaaacttgggagtgaaagctagagccaagg
ccatctaagtgacaggccaaagtgacacgaagcccacttcctgtgctccaaccatgag
ttccagcccaaaccaatggaaagtgatttacttgtcagggcccaaggacagtgca
gttctactccctccctcactaggagccaccttgggtgacagttgatcttaccactgt
aagtggtaaagggatggcctgggtcccaaccataatagggcggtggaaacggctcagg
agggtaacagcgtggattaggccacaagatggggcagatgatgtcatcagaagcatgtg

FIG 47A(IV)

112/126

accggtaggcaggtactaaacttctgggcaacctagtagccatgctatgcaggcaggt
agaggtagggcagtgctcatgtttggcatgatgatgtccacaaattcaggcttga
gagatgcgccaccacaaaggaagccgtccacgtcaggctggcttgccagctctttgca
ggttgctccagtcacagaaacctgtaccaggaacaagaagacagtttgggtcaggtctat
gatcagaacacttaagccccaccctctctgtgcaaggcagcctcagtcctttagccc
attccgtcttagctagagccaaagccactcacctccataaatgatccgggtgctctg
agccaccccatcatcgacattggatttcagccatccccggagcttctcgtgtacttcc
tgtgcctagaaggaggagcagagctactaagtaagctccttcctatctatcattcaa
ggagtaaaaaccactggttctcacatagagttgagttccagaaaaagccccggacca
gagagtggcaaggctccaatcccaccaggcttggaatgaacatttttggcaaaagtccac
tctccttggtgagtttgggggcccctctgtctctaaaaggggcttggtatgggctccatag
ctgtgtgagtcctgttaaagccggacaggctgaggagctctgggtagttacctgctgag
gggttgccgtcttgccagtcaccaatggcccacacaggttcataggccaggaccacctt
gctccagtccttcacattatctgtggggcagagaggagtagtaggaaaggagctga
cccccaagc

FIG 47A(V)

113/126

MGQTALARGSSSTPTSQALYSDFSPPEGLEELL SAPPD LVAQRHHGWNPKDCSENID
VKEGGLCFERRPVAQSTDGVRGKRGYSRGLHAW EISWPLEQRGTHAVVG VATALAPLQ
ADHYAALLGSNSESWGWDIGRGKLYHQSKGLEAPQYPAGPQGEQLV VPERLLVVLDME
EGTLGYSIGGTYLGPAFRGLKGR TLYPSVSAVWGQCQVRIRYMGERRRVEEPQSL LHL S
RLCVRHALGDTRLGOISTLPLPPAMKRYLLYK

FIG 47B

114/126

<u>FIG 48A (I)</u>
<u>FIG 48A (II)</u>
<u>FIG 48A (III)</u>
<u>FIG 48A (IV)</u>
<u>FIG 48A (V)</u>
<u>FIG 48A (VI)</u>

FIG 48A

115/126

gtactttcttataatccataaattttatttactattactacatgatacatatttta
taaaagtctttgtaacctccttaaggattcactgcttaattcctcagtgcttagcaca
atcattaaatgcgaaccagaactcttccaaatgtgttacatctataacctcattgga
ttctcactaccaaccccatgcaatagataactaatgtgatctctgtcttacagaggaag
aaacaggcacaggagggttcagtaatttgccaaaggtcatacacacactggccttcag
gtattcatgccccgggagtgctgggtccacagctggcatgtttggccattatatatt
gcctccttatagtgtcggcactcattaagcacattgacagctatgcttggtgagtga
tactatgtaccagctctgtgtacatgctttacctggattatttcaactgcacaaca
acctgtgaggtaactaccatcatgtctcctattttacataacagaaaactacagaaa
tctggggctgggcgtagtggtcctcatgccctgaaatcccagcactttgggagaccctgtc
tctaaaaaaatttttttgccggacgtgggtggctcacacctgtaatctcagcact
ttgggaggctaaaggcaggcagatcacaaaggtcaggagttctagaccagcctggccaac
atggcaaaaccctgtgtctactaaaaatacaaaaaatagctaggcgtggtggcagggtg
cctgtaatcccagctactcaggagggtgaggcaggagaatcccctgaacctgggagat
ggaggttacagagagccgagatcgtgccgctgcactccagcctgggcaacaagagcaa
gactctgtctcgaaaaaaataaaaaataaaaaataattttttaaaaaattagctg
gggtggttagcacatgcctgttagtcccagctacttgggaggctgaggtaggagatca
cttgagcccaggagggtcaaggctgcagtggtgggtgtgatggcgccactgcactctagcc

FIG 48A(I)

116/126

ttggtgacagcaagaccctgtctcaaaaaaaaaaagagaaatcgggcaacttccc
caagatcgcgagttaactagtggcatagtctcactcaaaactcgaagtcttaatcagg
aactctaccaaatgagatcaacggctcagtaattggattggcatccagtatgaagact
ggaccagcaggagaaactatgatgctgtacagcctagagcctgaagcagatttcacagc
ctcagagggtggcacaggctgactcacaacccggggcagaaaaggaccagccagaaac
agtgaccagaatcacagggaagtagaaatgggattcggcacaaatgaagccctcctt
gacccatgctccttaccctcaggggcagaggttagtgcctcaggcggctcaaaagg
tcttgacgggtggagaaacaccatccccagggtatcccgcgcgggtgatgccatcaaagc
gttaattctgagatgggctgcccgggtgcggactctgccgcagcaagagaagggtta
actgccccggccttcgccgtgggggcgggctcggggagggtcacagcccgggact
gagaccgagggttaaccgcccgggtgggctccacggggcgggcatgctctccgcg
gctgctgccgggtatagagcggtaactgcccaggaggggggggcccccacaggggcgt
ggcctcgagctgcacggccgtgggcggcgatgagagggttaagccccagagggccct
ggaggggcgggcgcgggacgggctcggcccaaggaggagagctgggggcggaagcgg
ccggcgggtctgcgcccctgcgcgccttcggcttcttccgcccggctccttcagaggccc
ggcgacctccagggctgggaagtcaaccgaggttcgggggcagcggcgagggtccgg
gcgagtaagggggatggtccatgctgagggcccaaatggggcgaaactcgcgagagtctc
tggcgacctggatcagatggggcgagggcagatgaaggggccagaggttgggggcag

FIG 48A(II)

117/126

cgaggaggagcgggcccgttggaacttggtgaaaggatgggtacctgggt
gacgagccccccaggattctgctcttcacgcccccttctccagctccctccag
gtcaatccaaactggagctcaactttcagaagagaagacgcccagcaagcctctt
cgggagtccttagctcctcacctccATGGGCCAGACAGCTCTGGCAGGGGGCAGCA
GCAGACCCCCACGCCACAGGCCCTGTACCCCTGACCTCTCCTGTCCCGAGGGCTTGA
AGAGCTGCTGTGCACCCCTCCTGACCTGGGGCCAGCGGCCACGGTTGGAAC
CCCAAAGACTGTTCAGAGAACATCGAGGTC AAGGAAGGAGGTTGTACTTTGAGCGGC
GGCCCGTGGCCAGAGCACTGATGGGGCCCCGGGTAAGAGGGGCTATTCAAGGGGCCT
GCACGCCCTGGGAGATCAGCTGGCCCCCTAGAGCAGAGGGGCACGCATGCCGTGTGGC
GTGGCCACGGCCCTCGCCCCGCTGCAGACTACCACTACGCGCGCTGTCTGGGCAGCA
ACAGCGAGTCGTGGGCTGGGACATCGGGCGGGGAAGCTGTACCATCAGAGCAAGGG
GCCCCGAGCCCCAGTATCCAGCGGGAACTCAGGGTGAGCAGCTGGAGGTGCCAGAG
AGACTGCTGGTGTCTGGACATGGAGGAGGGAACCTCTGGGCTACGCTATTGGGGCA
CCTACCTGGGGCCAGCATTCGCGGACTGAAGGGCAGGACCCCTCTATCCGGCAGTAAG
CGCTGTCTGGGGCCAGTGCCAGGTCCGCATCCGCTACCTGGGCGAAAGAGAGgtgag
gcctgggacagacgtggggagaacttctgtccctggtggcagtggtttgggatggaa
actctctgacaaagacagaggggatggaccttcacagcctgcctcaacctctgtt
cagtgtgggaaaggctaggggtcttcacagctgttatttaatttaacccaacagcaa

FIG 48A(III)

118/126

tagagtgaaacaggcttgagaaagcaactttctcaagttctcttgccagtaaatgg
tgaaccttcagaaatggaggagggaactgcagggatgagagaattcaggagatatcaac
ccctgagcaagaggtgcaaaagcgttaggtactgggtttgatgtacaggtccaaaaagaa
ggatgggcagagccagggtaccaggctgtataccggattccctgggctctaacctgtc
tctgtgccacatacctacttccttcctcagccacacctctggatggagacactggggc
cctgggcaccaggaggagagcagtgaggaggcaggcccttagggtggggcagcagg
ggaggagcctccccaggaaactgactgggtccagggttgaggctgctctctgcagttg
tgtgggctgtagagtgaggggccatccctcctcactcagccccagctcccaagcctc
tgagtgcaaaagcctggggcagctccaccactgtcagagccaccttggcctgtgttta
gagggccttagccagctctcaccctccagctctgactagggtgtgtgaaatcttatac
tgggaggcagaactccgggtatctcaaatcccccttcagccaggtgggcacactcg
aagcaggaaagcagaaaggcatctgagtaggaccccgtagtttgaggacatctggctg
gtggctgcacccatacttacattccccctcctctctcccagCGGAGCCACACTCCC
TTCTGCACCTGAGCCGCTGTGTGTCGCCACAACTGGGGGATACCCGGCTCGGCCA
GGTGCTGCCCTGCCCTTGCCCCCTGCCATGAAGCGCTACCTGCTCTACCAAGTGAgcc
ctgtgataccacagactgtgctgaggtcttgccaccacccctcccccttggggaggtgg
ggaggcactgctggcctagaccagctgctgaaagctggtgaggtgagccccctacccc
aaccacaagctctgcggaaatcaacagccccagagccacttgaggaggaggaagaaaggg

FIG 48A (IV)

119/126

agcggcggttaagggtatgacagctctgctacgcaaaacattttttcaagtaaaaata
gtaagagatgtgttatagaaacctgttcttgttttttttttcttgcaacaatga
tcatttatagctgcctcaaaaagggaagattatctgggcaagtcagtgaaaggcaga
caaccacaagacctagtgccagggttatccctcacatgggtgttcacatacacag
cacagaggcacgggcaccatgggagaggcagcactcctgccttctgaggggatcttg
gcctcacggtgtaagaaaggagaggatggtttcttctgcctcactagggcctagg
gaaccaggagcaaatcccaccgccttccatctctcagccaaggagaagccacctt
ggtgacgtttagttccaaccattatagtaagtggaaggattggcctgggtcccaac
cattacagggtgaagatataaacagtaaaagggaagatacagtttggtatgagggccacagg
aaggagcagatgacaccatcagaagcatatgcagggaaggcagttactgggcttct
gggtgcttagtccctggcttggcaggaaagggtagggaagatggatggggctcattgt
ttggcattgatgtccacgaattcgggcttgagggaagcaccaccacaaggaaagc
catccacatcaggctggctggccagctccttgcaagggtgccccagtcacagagcctgg
gaaggagcagaacaagggttggtcaagaaatgggatgagttgccccatccccacct
ccatgtccgagggtcagttcagttcctcagccccactccacctcagccgggaaccaaag
ccactcacctccataaatgatacgggtgctctgagccaccgcatcagagacgttggaac
ttcagccatcctcggagcttctcgtgtacttcttgggctagaacaagaagctggcct
aagtaagaccttttctgcctctcttaagaggaaaaatcactggcaccagtggaactta

FIG 48A(V)

120/126

gtgtgggtttctgactgagtcagagtagcagggtctctgatccaaagccaggcccttgact
ggatgcccttggacaagtcactgtctctgggttcaagggtctctgtgtctttgaaataa
gggttgcccccatgtgggctgtgtctgtccaaacctattgaggcaggctgggatgagg
gcagggtcctgggcccggttacctgttggggtgtgcagtcctgccagtaccaatgg
ccacacagggtcataggccaggacgaccttgctccagtccttcacgttatctgcagg
gcagagatacagatggaggggaagggtgaacaagaaagagctctccaggcaggttctcc
ggagtacgaagaacggtggcctactgccccctagtggaacattggggg

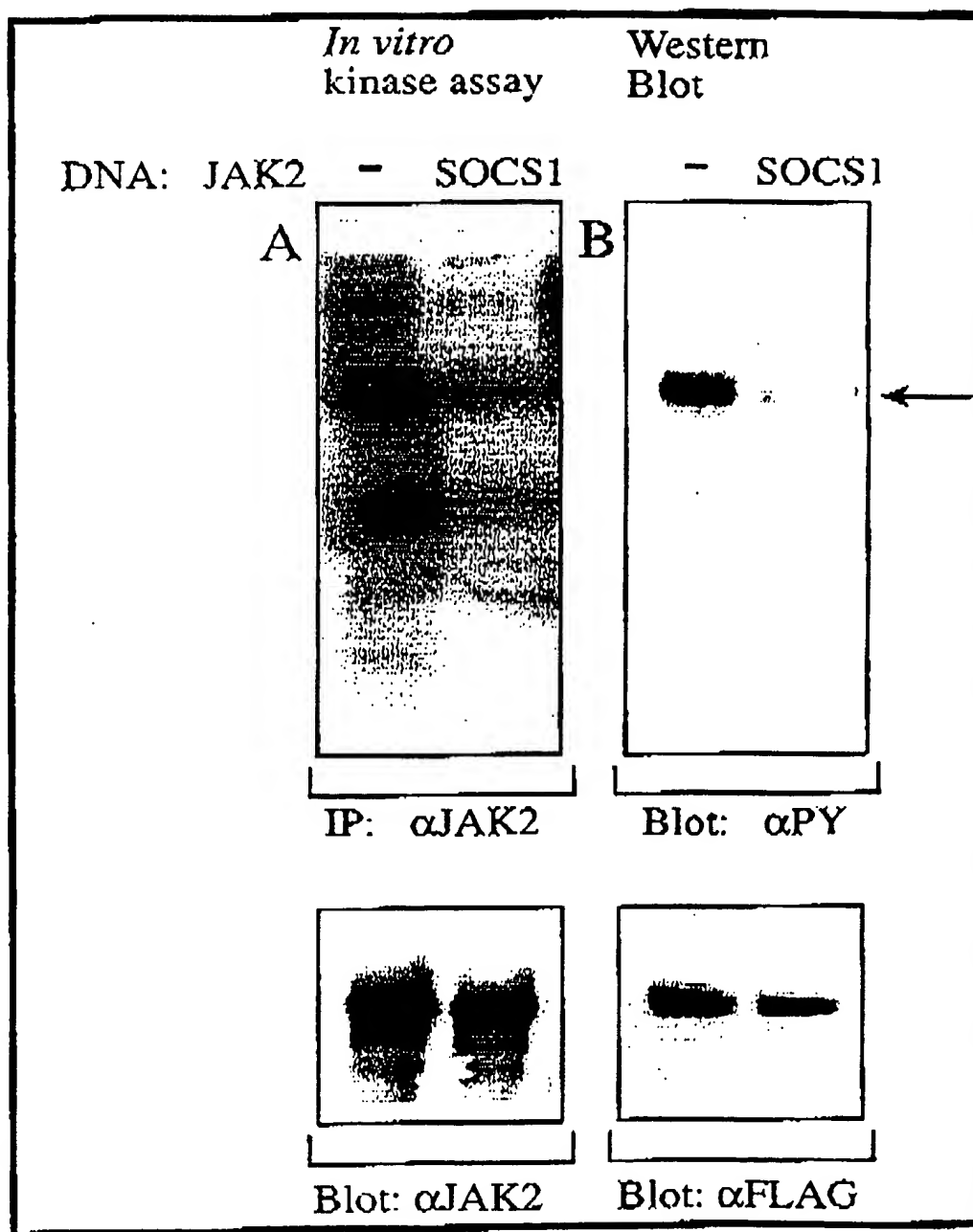
FIG 48A(VI)

121/126

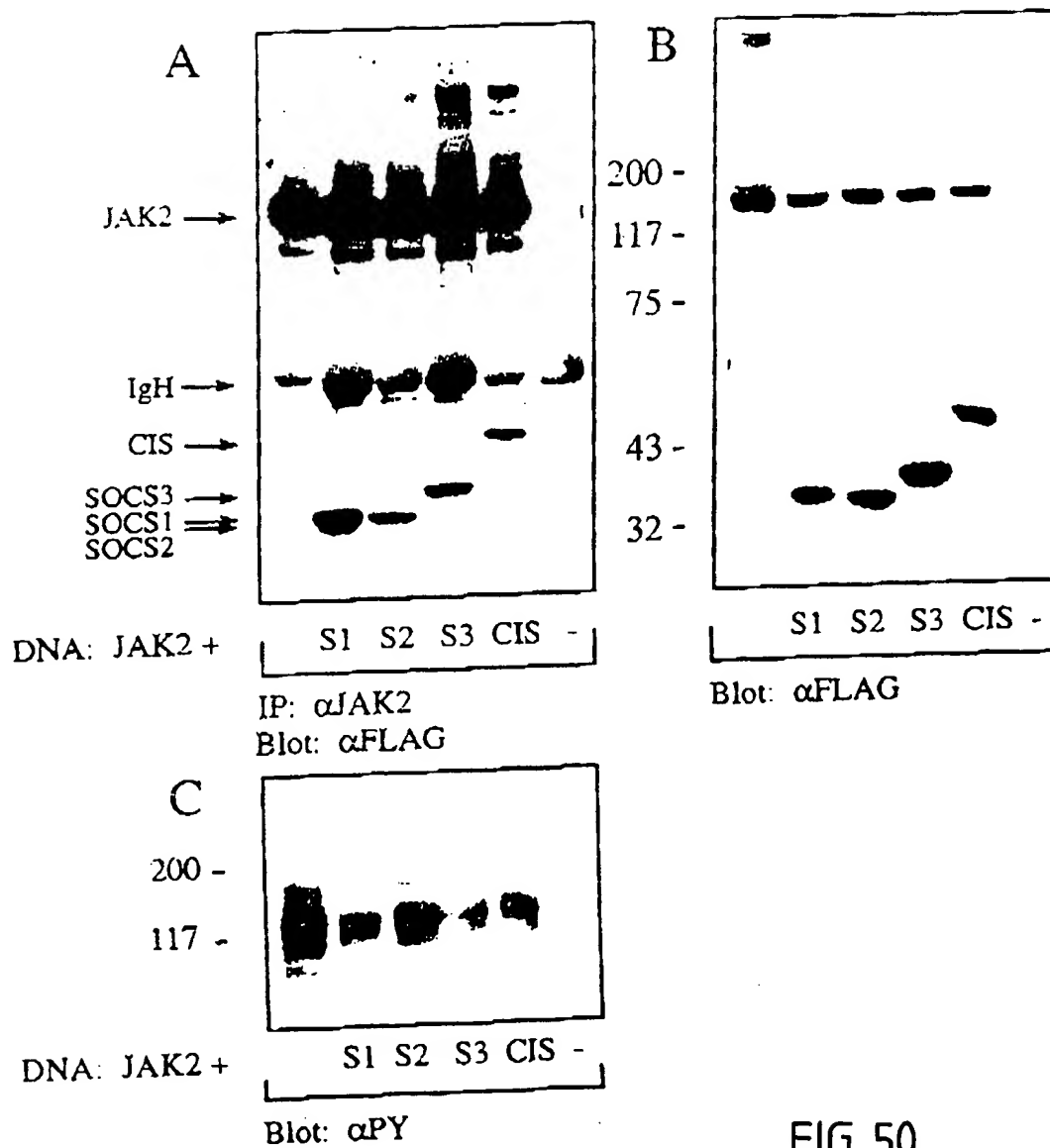
MGQTALAGSSSTPTPQALYPDDLSCPEGLEELLSAPPPDLGAQRRHGWNPKDCSENIE
VKEGGLYFERRPVAQSTDGARGKRGYSRGLHAWELISWPLEQQRGTHAVVGVATALAPLQ
TDHYAALLGSNSESNGWDIGRGKLYHQSKGPGAPQYPAGTQGEQLEVPERLLVVLDME
EGTLGYAIGGTYLGPFRGLKGRITLYPAVS AVWGQCQVRIRYLGERRAEPHSLHLISR
LCVRHNLGDTRLGOVSALPLPPAMKRYLLYQ

FIG 48B

122/126

FIG 49

123/126

FIG 50

124/126

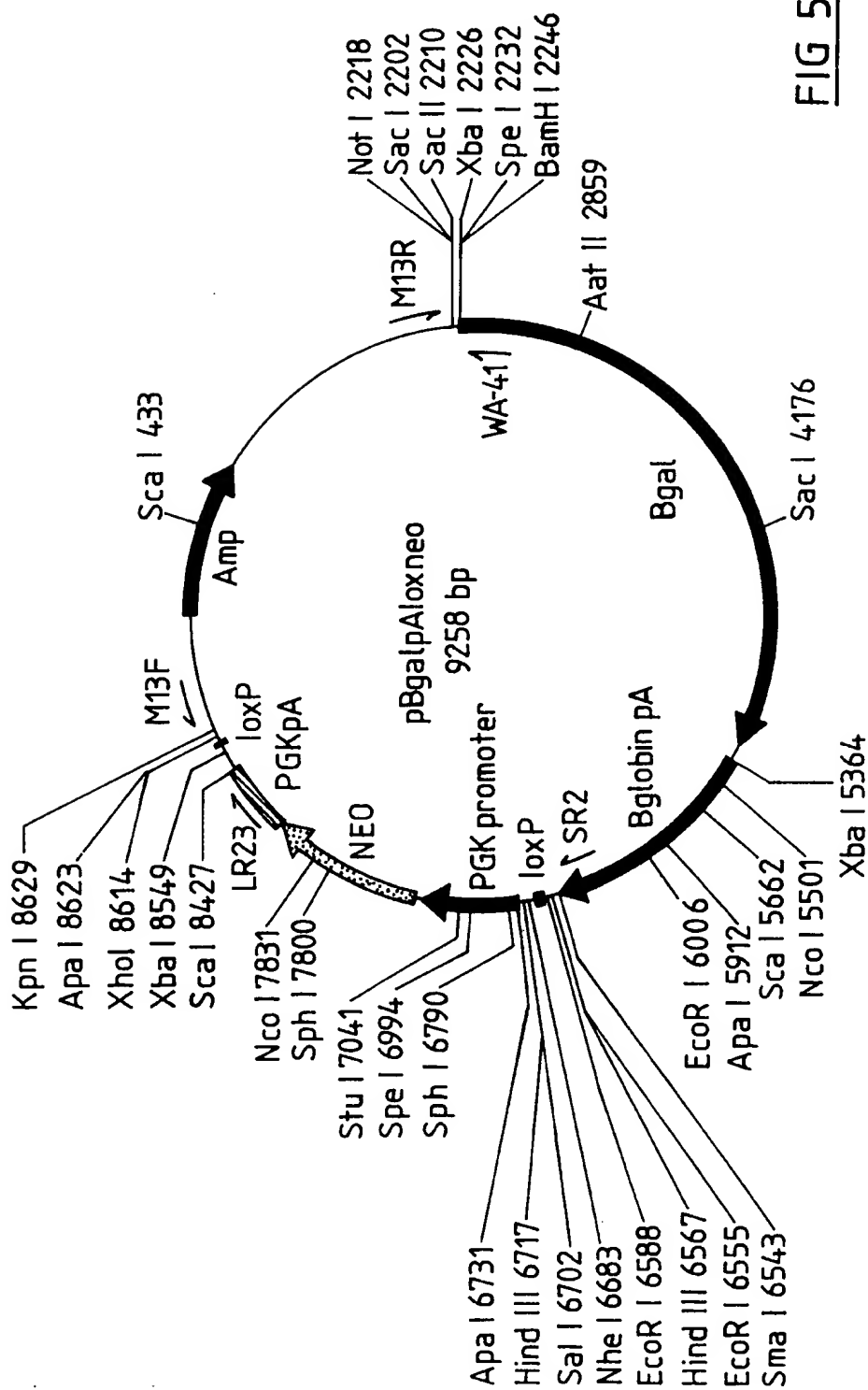


FIG 51

125/126

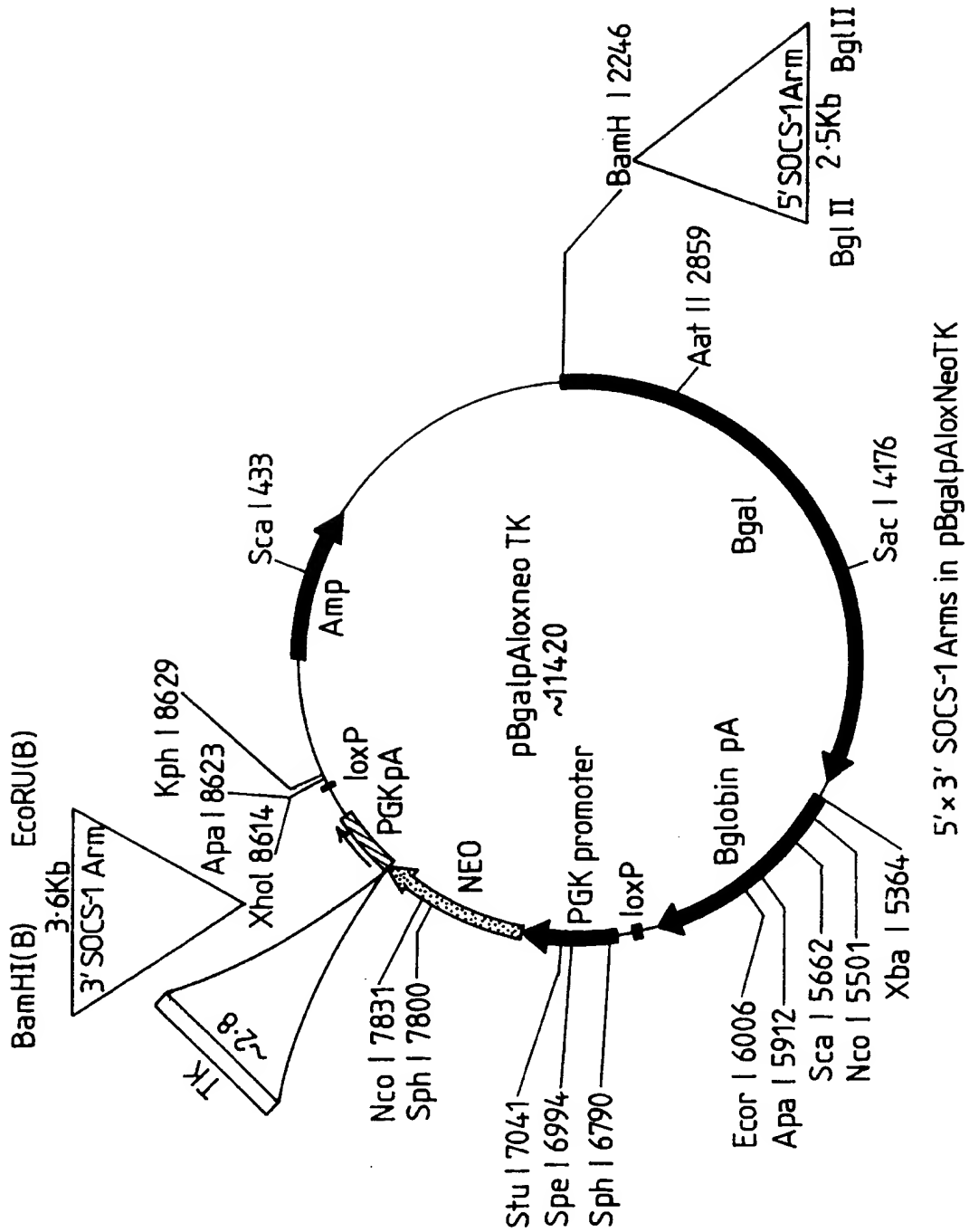


FIG 52

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 97/00729

A. CLASSIFICATION OF SUBJECT MATTER		
Int Cl ⁶ : CO7K 2/00		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) STN Peptide sub sequence search STN [LIVMAP]. [PTS]..[LIVMAP]..[LIVMAF YW] [CTS] [RKH]..[LIVMAP] {3}[LIVMAPGC TS].{1, ⁵⁰ }[LIVMAP]..[LIVMAP] P [LIVMAPG] [PN]. {1, ⁵⁰ } [LIVMAP]..[YF] [LIVMAP]		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	WO 96/39427 (Trustees of Dartmouth College) 12 December 1996 The whole document	1-40
X	<u>Yeast</u> vol 12 No 15 issued 1996 Delaveau, Th et al. "Analysis of a 23 kb region on the left arm of yeast chromosome IV" pages 1587-1592	1-40
X	<u>Science</u> vol 270 No 5234 issued 1995 Iabreit, Set al "Titins: giant proteins in charge of muscle ultrastructure and elasticity" pages 293-6	1-40
<input type="checkbox"/> Further documents are listed in the continuation of Box C <input checked="" type="checkbox"/> See patent family annex		
<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&" document member of the same patent family</p>		
Date of the actual completion of the international search 20 November 1997		Date of mailing of the international search report 12 DEC 1997
Name and mailing address of the ISA/AU AUSTRALIAN INDUSTRIAL PROPERTY ORGANISATION PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No.: (02) 6285 3929		Authorized officer K.F. PECK Telephone No.: (02) 6283 2263

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 97/00729

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<u>The EMBO Journal</u> Vol 14 No 12 issued 1995 Yoshimura, A et al "A novel cytokine - inducible gene CIS encodes and SH2 - containing protein that binds to tyrosine - phosphorylated interleukin 3 and erythropoietin receptors" pages 2816-26	1-40
X	AU, A, 27924/95 (Flügge, U.I.) 17 August 1995 The whole document, particularly pages 29-32	1-40
X	<u>Biochemistry</u> vol 34 No 8 issued 1995 Weber, A et al "The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells." pages 2621-7	1-40
X	<u>Journal of Bacteriology</u> Vol 176 No 24 issued 1994 Iwai, A et al "Molecular cloning and expression of an isomaltose-dextranase gene from <i>Arthrobacter globiformis</i> T6" pages 7730-4.	1-40
X	<u>Nucleic Acids Research</u> Vol 122 No 11 issued 1994 Althoff, S et al "Molecular evolution of SRP cycle components: functional implications" pages 1933-47	1-40
X	<u>Nature</u> vol 368 No 6466 issued 1994 Wilson, R et al "2.2 Mb of contiguous nucleotide sequence from chromosome III of <i>C. elegans</i> " pages 32-8.	1-40
X	<u>The EMBO Journal</u> Vol 11 No 5 issued 1992 Labeit, S et al "Towards a molecular understanding of Titin" pages 1711-16	1-40
X	<u>Advances in Biophysics</u> Vol 33 (Muscle Elastic Proteins) issued 1996 Kolmerer, B et al "A systematic search of the data bases for sequences homologous to titin/connectin" pages 3-11	1-40
X	<u>Microbiology</u> Vol 142 no 8 issued 1996 Yoneyama, H "Protein C (OprC) of the outer membrane of <i>Pseudomonas aeruginosa</i> is a copper-regulated channel protein" pages 2137-2144.	1-40
X	<u>Journal of Bacteriology</u> Vol 178 No 15 issued 1996 Limberger, R et al. "Organisation, transcription and expression 'of the 5' region of the fla operon of <i>Treponema phagedenis</i> and <i>Treponema pallidum</i> " pages 4628-4634.	1-40
X	<u>The Journal of cell biology</u> Vol 133 No 6 issued 1996 Goodson, H. V et al "Synthetic lethality screen identifies a novel yeast myosin I gene (MYO5): myosin I protein are required for polarisation of the actin cytoskeleton" pages 1277-1291	1-40
X	<u>Genes and Development</u> Vol 9 No 24 issued 1995 Herrscher, R F et al "The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family" pages 3067-82	1-40

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.
PCT/AU 97/00729

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report			Patent Family Member				
AU, A	27924/95	CA	2192849	DE	4420782	EP	765393
		HY	9603441	WO	95/34654		